GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:19:30 ; Search time 10 Seconds
(without alignments)
2836.981 Million cell updates/sec

Title: US-09-518-842-1_COPY_76_417
Perfect score: 635
Sequence: 1 GCAGAGCTGAGGGGATGTGG.....CTTCAGTTAAATTATGTACA 342

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09518842/runat_14062003_175810_12259/app_query.fasta_1.519
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09518842_@CGN_1_1_26_@runat_14062003_175810_12259 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	619	97.5	139	1	INL4_HUMAN	Q14641 homo sapien
	2	141	22.2	182	1	RELX_PIG	P01348 sus scrofa
	3	116.5	18.3	185	1	RELX_MOUSE	P47932 mus musculu
	4	111	17.5	178	1	RELH_RABIT	P51456 oryctolagus
	5	106.5	16.8	182	1	RELX_HORSE	P22969 equus cabal
	6	105.5	16.6	166	1	REL2_PANTR	P51455 pan troglod
	7	105.5	16.6	185	1	REL2_HUMAN	P04090 homo sapien
	8	103.5	16.3	185	1	REL1_HUMAN	P04808 homo sapien
	9	103.5	16.3	185	1	RELX_MACMU	P19884 macaca mula
	10	100.5	15.8	166	1	REL1_PANTR	P51454 pan troglod
	11	97.5	15.4	186	1	RELX_RAT	P01347 rattus norv
	12	76.5	12.0	160	1	RELX_CAVPO	P51453 cavia porce
	13	75	11.8	836	1	DPOL_HPBDU	P03162 duck hepati
	14	74.5	11.7	1531	1	PMPD_CHLTR	O84818 chlamydia t
	15	73.5	11.6	899	1	Y188_HUMAN	Q14693 homo sapien
c	16	73	12.2	857	1	AR56_CANAL	P78586 c arg5,6 pr
c	17	73	12.2	993	1	RPN1_YEAST	P38764 saccharomyc
	18	73	11.5	1395	1	SP41_YEAST	P38904 saccharomyc
	19	72.5	11.4	292	1	YHFR_SALTI	Q8z4m8 salmonella
c	20	72.5	12.1	1744	1	CO4_HUMAN	P01028 homo sapien
	21	72	11.3	3358	1	PGCV_MOUSE	Q62059 mus musculu
	22	71.5	11.3	292	1	YFHR_SALTY	Q8zn39 salmonella
	23	70.5	11.1	271	1	VG18_BPP22	P03687 bacteriopha
	24	70.5	11.1	957	1	UVRA_BACSU	O34863 bacillus su
	25	70	11.0	283	1	PANC_SCHPO	Q09673 schizosacch
	26	69.5	10.9	645	1	SYT_BACHD	Q9k866 bacillus ha
	27	69	10.9	177	1	RELX_MESAU	Q64171 mesocricetu
c	28	69	11.5	274	1	RRPP_VSVJM	P04878 vesicular s
c	29	69	11.5	943	1	BLI4_CAEEL	P51559 caenorhabdi
	30	69	10.9	3396	1	PGCV_HUMAN	P13611 homo sapien
c	31	68	11.3	503	1	CP3P_MOUSE	O09158 mus musculu
	32	67.5	10.6	404	1	RTP1_TRYBG	P15593 trypanosoma
	33	67	10.6	313	1	SDC1_RAT	P26260 rattus norv
c	34	67	11.2	449	1	YKE6_YEAST	P36091 saccharomyc
	35	67	10.6	629	1	DXS_DEIRA	Q9rub5 deinococcus
	36	67	10.6	3828	1	TRX_DROVI	Q24742 drosophila
c	37	66.5	11.1	123	1	ELI2_PHYCR	P41803 phytophthor
	38	66.5	10.5	295	1	YS86_MYCTU	Q10831 mycobacteri
	39	66.5	10.5	979	1	T111_SCHPO	Q09813 schizosacch
c	40	66	11.0	897	1	CYRB_HUMAN	P32927 homo sapien
	41	66	10.4	1025	1	CR2_MOUSE	P19070 mus musculu
c	42	66	11.0	3695	1	LMA5_HUMAN	O15230 homo sapien
	43	65.5	10.3	443	1	NRTA_SYNP7	P38043 synechococc
	44	65.5	10.3	529	1	GAG_MLVDU	P23090 duplan muri
c	45	65.5	10.9	564	1	GAF2_SCHPO	Q10134 schizosacch

ALIGNMENTS

RESULT 1
INL4_HUMAN
ID INL4_HUMAN STANDARD; PRT; 139 AA.
AC Q14641;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Early placenta insulin-like peptide precursor (EPIL) (Placentin)
DE (Insulin-like peptide 4).
GN INSL4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96115599; PubMed=8666396;
RA Chassin D., Laurent A., Janneau J.-L., Berger R., Bellet D.;
RT "Cloning of a new member of the insulin gene superfamily (INSL4)
RT expressed in human placenta.";
RL Genomics 29:465-470(1995).
RN [2]
RP PROCESSING.
RX MEDLINE=97430657; PubMed=9284764;
RA Bellet D., Lavaissiere L., Mock P., Laurent A., Sabourin J.C.,
RA Bedossa P., Le Bouteiller P., Frydman R., Troalen F., Bidart J.M.;
RT "Identification of pro-EPIL and EPIL peptides translated from
RT insulin-like 4 (INSL4) mRNA in human placenta.";
RL J. Clin. Endocrinol. Metab. 82:3169-3172(1997).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=98411035; PubMed=9740319;
RA Laurent A., Rouillac C., Delezoide A.L., Giovangrandi Y., Vekemans M.,
RA Bellet D., Abitbol M., Vidaud M.;
RT "Insulin-like 4 (INSL4) gene expression in human embryonic and
RT trophoblastic tissues.";
RL Mol. Reprod. Dev. 51:123-129(1998).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN TROPHOBLAST DEVELOPMENT
CC AND IN THE REGULATION OF BONE FORMATION.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, UTERUS AND IN FETAL
CC PERICHONDIRUM.

```
CC   -!- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN THE EARLY PLACENTA.
CC       EXPRESSION OF EPIL PEPTIDES IN THE VILLOUS CYTOTROPHOBLAST IS
CC       DIFFERENT FROM THAT DISPLAYED BY THE SYNCYTIOTROPHOBLAST. IN FETAL
CC       TISSUES IT WAS IDENTIFIED IN THE PERICHONDRIUM OF ALL FOUR LIMBS,
CC       VERTEBRAE, AND RIBS. IT WAS ABUNDANT IN INTERBONE LIGAMENTS.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34838; AAB08516.1; -.
DR   Genew; HGNC:6087; INSL4.
DR   MIM; 600910; -.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    139       EARLY PLACENTA INSULIN-LIKE PEPTIDE.
FT   CHAIN        18     58       EARLY PLACENTA INSULIN-LIKE PEPTIDE B
FT                                CHAIN.
FT   PROPEP       59    109       C PEPTIDE.
FT   CHAIN       110    139       EARLY PLACENTA INSULIN-LIKE PEPTIDE A
FT                                CHAIN.
FT   DISULFID     31    125       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     43    138       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    124    129       BY SIMILARITY.
SQ   SEQUENCE   139 AA;  15445 MW;  47FB61F6F86C1342 CRC64;

Alignment Scores:
Pred. No.:              2.69e-60        Length:         139
Score:                  619.00          Matches:        114
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            97.48%          Indels:         0
DB:                     1               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x INL4_HUMAN (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              ||||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 2
RELX_PIG
ID   RELX_PIG       STANDARD;      PRT;   182 AA.
AC   P01348;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor.
GN   RLN.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87308187; PubMed=2442155;
RA   Haley J., Crawford R., Hudson P., Scanlon D., Tregear G., Shine J.,
RA   Niall H.;
RT   "Porcine relaxin. Gene structure and expression.";
RL   J. Biol. Chem. 262:11940-11946(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83157118; PubMed=6897721;
RA   Haley J., Hudson P., Scanlon D., John M., Cronk M., Shine J.,
RA   Tregear G., Niall H.;
RT   "Porcine relaxin: molecular cloning and cDNA structure.";
RL   DNA 1:155-162(1982).
RN   [3]
RP   PRELIMINARY SEQUENCE OF 25-54 AND 161-182.
RX   MEDLINE=77213067; PubMed=876374;
RA   James R., Niall H., Kwok S., Bryant-Greenwood G.;
RT   "Primary structure of porcine relaxin: homology with insulin and
RT   related growth factors.";
RL   Nature 267:544-546(1977).
RN   [4]
RP   SEQUENCE OF 25-51.
RX   MEDLINE=77157271; PubMed=851452;
RA   Schwabe C., McDonald J.K., Steinetz B.G.;
RT   "Primary structure of the B-chain of porcine relaxin.";
RL   Biochem. Biophys. Res. Commun. 75:503-510(1977).
RN   [5]
RP   SEQUENCE OF 25.
RX   MEDLINE=77134136; PubMed=843375;
RA   Schwabe C., McDonald J.K.;
RT   "Demonstration of a pyroglutamyl residue at the N terminus of the B-
RT   chain of porcine relaxin.";
RL   Biochem. Biophys. Res. Commun. 74:1501-1504(1977).
RN   [6]
RP   SEQUENCE OF 161-182.
RX   MEDLINE=76231539; PubMed=938497;
RA   Schwabe C., McDonald J.K., Steinetz B.G.;
RT   "Primary structure of the A chain of porcine relaxin.";
RL   Biochem. Biophys. Res. Commun. 70:397-405(1976).
RN   [7]
RP   DISULFIDE BONDS.
RX   MEDLINE=77236040; PubMed=887933;
RA   Schwabe C., McDonald J.K.;
RT   "Relaxin: a disulfide homolog of insulin.";
RL   Science 197:914-915(1977).
RN   [8]
RP   3D-STRUCTURE MODELING.
RX   MEDLINE=78092399; PubMed=622170;
RA   Isaacs N.W., James R., Niall H., Bryant-Greenwood G., Dodson G.G.,
RA   Evans A., North A.C.T.;
RT   "Relaxin and its structural relationship to insulin.";
RL   Nature 271:278-281(1978).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN
CC       TO PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K01088; AAA31114.1; -.
DR   EMBL; J02792; AAA31115.1; -.
DR   EMBL; A16593; CAA01295.1; -.
DR   EMBL; A06852; CAA00600.1; -.
DR   PIR; A01615; RXPG.
DR   PIR; A29796; A29796.
DR   PDB; 1RLX; 15-OCT-94.
DR   PDB; 2RLX; 15-OCT-94.
DR   PDB; 3RLX; 15-OCT-94.
DR   PDB; 4RLX; 15-OCT-94.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal; 3D-structure.
FT   SIGNAL        1     24
FT   CHAIN        25     56       RELAXIN B CHAIN.
FT   PROPEP       57    154       CONNECTING PEPTIDE.
FT   CHAIN       161    182       RELAXIN A CHAIN.
FT   MOD_RES      25     25       PYRROLIDONE CARBOXYLIC ACID.
FT   DISULFID     34    169       INTERCHAIN.
FT   DISULFID     46    182       INTERCHAIN.
FT   DISULFID    168    173
FT   CONFLICT     47     47       G -> GVWS (IN REF. 4).
FT   CONFLICT     51     54       WGRT -> TWGR (IN REF. 3).
FT   CONFLICT    116    116       S -> L (IN REF. 1).
FT   CONFLICT    170    170       Q -> E (IN REF. 6).
SQ   SEQUENCE   182 AA;  20818 MW;  20736EB089F13AB4 CRC64;

Alignment Scores:
Pred. No.:                1.42e-07       Length:          182
Score:                    141.00         Matches:         41
Percent Similarity:       40.00%         Conservative:    7
Best Local Similarity:    34.17%         Mismatches:      32
Query Match:              22.20%         Indels:          40
DB:                       1              Gaps:            2

US-09-518-842-1_COPY_76_417 (1-342) x RELX_PIG (1-182)

Qy        100 CTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCC 159
              ||||||:::|||   |||   |||   :::   :::|||      ||||||   :::
Db         63 LeuGluThrGlyProProAlaGluThrMetProSerSerIleThrLysAspAlaGluIle 82

Qy        160 TTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 219
              |||            ||||||:::||||||||      |||||||||      ||||||
Db         83 LeuLysMetMetLeuGluPheValProAsnLeuProGlnGluLeuLysAlaThrLeuSer 102

Qy        220 GAAGGGCAGCCATCATTG---------------------------------------- 237
              |||   ||||||||||||
Db        103 GluArgGlnProSerLeuArgGluLeuGlnGlnSerAlaSerLysAspSerAsnLeuAsn 122

Qy        238 ------------AAGAAAATAATACTT------------------------------- 252
                         ||||||||||||||
Db        123 PheGluGluPheLysLysIleIleLeuAsnArgGlnAsnGluAlaGluAspLysSerLeu 142

Qy        253 --------------------------------TCCCGCAAAAAGAGAAGTGGACGTCAC 279
                                              ||||||||||||||||      |||
Db        143 LeuGluLeuLysAsnLeuGlyLeuAspLysHisSerArgLysLysArgLeuPheArgMet 162

Qy        280 AGATTTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                          ||||||:::|||   |||                 :::||||||
Db        163 ThrLeuSerGluLysCysCysGlnValGlyCysIleArgLysAspIleAlaArgLeuCys 182

RESULT 3
RELX_MOUSE
ID   RELX_MOUSE     STANDARD;      PRT;   185 AA.
AC   P47932;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor.
GN   RLN OR RLX.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss; TISSUE=Ovary;
RX   MEDLINE=93199663; PubMed=8452637;
RA   Evans B.A., John M., Fowler K.J., Summers R.J., Cronk M.,
RA   Shine J., Tregear G.W.;
RT   "The mouse relaxin gene: nucleotide sequence and expression.";
RL   J. Mol. Endocrinol. 10:15-23(1993).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN
CC       TO PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z27088; CAA81611.1; -.
DR   MGD; MGI:97931; Rln.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; FALSE_NEG.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23     57       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       58    156       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       161    185       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     36    171       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     48    185       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    170    175       BY SIMILARITY.
SQ   SEQUENCE   185 AA;  20570 MW;  2B9E753E8B85087B CRC64;

Alignment Scores:
Pred. No.:                7.13e-05       Length:          185
Score:                    116.50         Matches:         38
Percent Similarity:       35.76%         Conservative:    16
Best Local Similarity:    25.17%         Mismatches:      56
Query Match:              18.35%         Indels:          41
DB:                       1              Gaps:            2

US-09-518-842-1_COPY_76_417 (1-342) x RELX_MOUSE (1-185)

Qy          7 CTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAG 66
              :::|||   ||||||      :::   :::   |||:::      |||
Db         33 IleArgMetCysGlyArgGluTyrAlaArgGluLeuIleLysIleCysGlyAlaSerVal 52

Qy         67 AAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATG 126
                        :::               |||   :::   :::      |||:::
Db         53 GlyArgLeuAlaLeuSerGlnGluGluProAlaLeuLeuAlaArgGlnAlaThrGluVal 72

Qy        127 GTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCT 186
              |||   :::      ||||||||      |||||||||   |||   ||||||   :::   :::|||
Db         73 ValProSerPheIleAsnLysAspAlaGluProPheAspThrThrLeuLysCysLeuPro 92

Qy        187 AATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATA 246
              |||||||||   |||||||||      |||||||||   |||   ||||||   ::::::
Db         93 AsnLeuSerGluGluLeuLysAlaValLeuSerGluAlaGlnAlaSerLeuProGluLeu 112

Qy        247 ------------ATACTT------------------------------------------ 252
                        :::|||
Db        113 GlnHisAlaProValLeuSerAspSerValValSerLeuGluGlyPheLysLysThrLeu 132
```

```
Qy        252 ------------------------------------------------------------ 252

Db        133 HisAspArgLeuGlyGluAlaGluAspGlySerProProGlyLeuLysTyrLeuGlnSer 152

Qy        253 ---------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAA 303
                      ||||||||||||||||                              ||||||
Db        153 AspThrHisSerArgLysLysArgGluSerGlyGlyLeuMetSerGlnGlnCysCysHis 172

Qy        304 GTAATTTGTGACGATGGAACTTCAGTTAAATTA 336
                 |||   |||       :::     ||||||
Db        173 ValGlyCysSerArgArgSerIleAlaLysLeu 183
```

RESULT 4
RELH_RABIT
ID   RELH_RABIT   STANDARD;   PRT;   178 AA.
AC   P51456;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Relaxin-like protein SQ10 precursor.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Tracheobronchial epithelium;
RX   MEDLINE=93002619; PubMed=1339318;
RA   Jetten A.M., Bernacki S.H., Floyd E.E., Saunders N.A., Pieniazek J.,
RA   Lotan R.;
RT   "Expression of a preprorelaxin-like gene during squamous
RT   differentiation of rabbit tracheobronchial epithelial cells and its
RT   suppression by retinoic acid.";
RL   Cell Growth Differ. 3:549-556(1992).
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC   DISULFIDE BONDS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- INDUCTION: DURING SQUAMOUS CELL DIFFERENTIATION. REPRESSED BY
CC   RETINOIC ACID.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; S45940; AAB23648.1; -.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL   1   20   POTENTIAL.
FT   CHAIN   21   52   RELAXIN-LIKE PROTEIN SQ10 B CHAIN
FT   (POTENTIAL).
FT   PROPEP   53   150   CONNECTING PEPTIDE (POTENTIAL).
FT   CHAIN   155   178   RELAXIN-LIKE PROTEIN SQ10 A CHAIN
FT   (POTENTIAL).
FT   DISULFID   34   165   INTERCHAIN (BY SIMILARITY).
FT   DISULFID   46   178   INTERCHAIN (BY SIMILARITY).
FT   DISULFID   164   169   BY SIMILARITY.
SQ   SEQUENCE   178 AA; 20294 MW; F6A54D98A6B53211 CRC64;

Alignment Scores:
Pred. No.:   0.000286   Length: 178
Score: 111.00   Matches: 39
Percent Similarity: 32.81%   Conservative: 3
Best Local Similarity: 30.47%   Mismatches: 42
Query Match: 17.48%   Indels: 44

```
DB:                    1             Gaps:          3

US-09-518-842-1_COPY_76_417 (1-342) x RELH_RABIT (1-178)

Qy         37 CACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGG 96
              ||||||          |||   ||||||      |||
Db         50 HisLeuGluArgGluSerProSerProGluAsnProPhe---------------------- 62

Qy         97 CTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAA 156
                 |||   ||||||         |||   |||   :::|||      ||||||   :::
Db         63 ---LeuSerSerGlyProAlaAlaGluThrValProSerSerIleLysLysAspAlaGlu 81

Qy        157 GCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTG 216
                    |||       |||   ||||||||||||      ||||||         |||
Db         82 AsnAlaAsnThrMetLeuGluSerIleProAsnLeuProGlnGluLeuThrAlaThrLeu 101

Qy        217 TCTGAAGGGCAGCCATCA------------------------------------------ 234
                 |||   |||||||||
Db        102 PheGluLysGlnProSerLysLeuTyrLeuGlnTyrLeuProThrLeuLysLysSerAsn 121

Qy        235 --------------TTGAAGAAAATAATA---------------------------- 249
                            ||||||||||||
Db        122 ValSerPheGluGluPheLysLysIleIleGlnAsnIleGlnArgGlyValGlnGlySer 141

Qy        250 --------------------CTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                   ||||||||||||||||
Db        142 SerAlaSerGluSerAsnThrPheSerArgLysLysArgGlnPheSerGluSerLeuPro 161

Qy        289 CCATTCTGTTGTGAAGTAATTTGT 312
                    |||||:::      |||
Db        162 GluGluCysCysLysTyrGlyCys 169
```

RESULT 5
RELX_HORSE
ID   RELX_HORSE   STANDARD;   PRT;   182 AA.
AC   P22969; Q28907;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor (RXN).
GN   RLN.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Hokkaido; TISSUE=Placenta;
RA   Min K., Shiota K., Ogawa T.;
RT   "Molecular cloning of equine preprorelaxin cDNA.";
RL   J. Reprod. Dev. 42:171-178(1996).
RN   [2]
RP   SEQUENCE OF 32-174 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=95359320; PubMed=7543295;
RA   Klonisch T., Ryan P.L., Yamashiro S., Porter D.G.;
RT   "Partial complementary deoxyribonucleic acid cloning of equine relaxin
RT   messenger ribonucleic acid, and its localization within the equine
RT   placenta.";
RL   Biol. Reprod. 52:1307-1315(1995).
RN   [3]
RP   SEQUENCE OF 26-53 AND 163-182.
RC   TISSUE=Placenta;
RX   MEDLINE=91275796; PubMed=2055195;
RA   Stewart D.R., Nevins B., Hadas E., Vandlen R.;
RT   "Affinity purification and sequence determination of equine relaxin.";
RL   Endocrinology 129:375-383(1991).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN
CC   TO PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC   DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB000201; BAA19069.1; -.
DR   EMBL; S78800; AAB35036.1; -.
DR   PIR; A49739; A49739.
DR   PIR; B49739; B49739.
DR   HSSP; P01348; 1RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     25
FT   CHAIN        26     53       RELAXIN B CHAIN.
FT   PROPEP       54    156       CONNECTING PEPTIDE.
FT   CHAIN       161    182       RELAXIN A CHAIN.
FT   DISULFID     35    169       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     47    182       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    168    173       BY SIMILARITY.
FT   CONFLICT     66     66       A -> V (IN REF. 2).
FT   CONFLICT    133    133       L -> Q (IN REF. 2).
SQ   SEQUENCE   182 AA;  20721 MW;  E5C9414303A838B8 CRC64;

Alignment Scores:
Pred. No.:              0.000901        Length:         182
Score:                  106.50          Matches:        30
Percent Similarity:     52.94%          Conservative:   15
Best Local Similarity:  35.29%          Mismatches:     33
Query Match:            16.77%          Indels:         7
DB:                     1               Gaps:           3

US-09-518-842-1_COPY_76_417 (1-342) x RELX_HORSE (1-182)

Qy          7 CTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCT--- 63
              ::::::   ||||||          :::      :::      |||
Db         32 IleLysAlaCysGlyArgGluLeuAlaArgLeuArgIleGluIleCysGlySerLeuSer 51

Qy         64 ---GAGAAGACATTC------ACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGT 114
                 :::|||||              ||||||         ||||||:::|||:::
Db         52 TrpLysLysThrValLeuArgLeuGluGluProGly---------LeuGluAlaGlyGln 68

Qy        115 CCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
              |||   |||:::||||||:::|||   :::||||||   :::||||||   |||
Db         69 ProValGluIleValSerSerSerIleSerLysAspAlaGluAlaLeuAsnThrLysLeu 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
                       ||||||      |||   |||      |||||||||   |||||||||
Db         89 GlyLeuAsnSerAsnLeuProLysGluGlnLysAlaThrLeuSerGluArgGlnProSer 108

Qy        235 TTGAAGAAAATAATA 249
              :::::::::::::
Db        109 TrpArgGluLeuLeu 113

RESULT 6
REL2_PANTR
ID   REL2_PANTR     STANDARD;      PRT;   166 AA.
AC   P51455; P79267;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin H2 precursor (Fragment).
GN   RNL2 OR RLX2.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC   TISSUE=Placenta;
RX   MEDLINE=94238260; PubMed=8182365;
RA   Evans B.A., Fu P., Tregear G.W.;
RT   "Characterization of two relaxin genes in the chimpanzee.";
RL   J. Endocrinol. 140:385-392(1994).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM 2), AND TISSUE SPECIFICITY.
RC   TISSUE=Placenta, and Ovary;
RX   MEDLINE=96328899; PubMed=8735594;
RA   Gunnersen J.M., Fu P., Roche P.J., Tregear G.W.;
RT   "Expression of human relaxin genes: characterization of a novel
RT   alternatively-spliced human relaxin mRNA species.";
RL   Mol. Cell. Endocrinol. 118:85-94(1996).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE CORPUS LUTEUM OF PREGNANCY
CC       AND IN THE PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z27245; CAA81758.1; -.
DR   EMBL; S83209; AAD14430.1; -.
DR   HSSP; P04090; 6RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Multigene family; Signal; Polymorphism;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   SIGNAL       <1      5       POTENTIAL.
FT   CHAIN         6     34       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       37    138       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       143    166       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     16    153       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     28    166       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    152    157       BY SIMILARITY.
FT   VARSPLIC     52     84       EIVPSFINKDTETINMMSEFVANLPQELKLTLS -> DFIQ
FT                                TVSLGISPDGGKALRTGSCFTREFLGALS (IN ISOFORM
FT                                2).
FT   VARSPLIC     85    166       MISSING (IN ISOFORM 2).
FT   VARIANT      30     34       KSTWS -> MSTLG (IN SOME ALLELES).
SQ   SEQUENCE   166 AA;  18760 MW;  220851E3134CEBDB CRC64;

Alignment Scores:
Pred. No.:              0.00114         Length:         166
Score:                  105.50          Matches:        43
Percent Similarity:     35.67%          Conservative:   13
Best Local Similarity:  27.39%          Mismatches:     44
Query Match:            16.61%          Indels:         57
DB:                     1               Gaps:           6

US-09-518-842-1_COPY_76_417 (1-342) x REL2_PANTR (1-166)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         23 AlaGlnIleAlaIleCysGlyLysSerThrTrpSerLysArgSerLeuSerGlnGluAsp 42
```

```
Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::               ||||||                           ||||||
Db         43 AlaProGln--------------ThrPro-----------------------ArgPro 49

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::      |||||||||   :::   :::         |||
Db         50 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 69

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
                 ||||||:::   ||||||      |||||||||      |||||||||   ||||||:::
Db         70 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 89

Qy        235 TTG------------------------------------------------------- 237
                 |||
Db         90 LeuProGlnLeuGlnGlnTyrValProValLeuLysAspSerSerLeuLeuPheGluGlu 109

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                 ||||||:::|||
Db        110 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 129

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||||||||||
Db        130 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 149

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                 ||||||   |||   |||        :::      :::   |||
Db        150 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 166
```

RESULT 7
REL2_HUMAN
ID   REL2_HUMAN   STANDARD;   PRT;   185 AA.
AC   P04090; Q9UCX3; Q99936;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Prorelaxin H2 precursor.
GN   RLN2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RX   MEDLINE=85051298; PubMed=6548702;
RA   Hudson P., John M., Crawford R., Haralambidis J., Scanlon D.,
RA   Gorman J., Tregear G., Shine J., Niall H.;
RT   "Relaxin gene expression in human ovaries and the predicted structure
RT   of a human preprorelaxin by analysis of cDNA clones.";
RL   EMBO J. 3:2333-2339(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA   Sehra H.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE
RP   SPECIFICITY.
RC   TISSUE=Prostate;
RX   MEDLINE=96328899; PubMed=8735594;
RA   Gunnersen J.M., Fu P., Roche P.J., Tregear G.W.;
RT   "Expression of human relaxin genes: characterization of a novel
RT   alternatively-spliced human relaxin mRNA species.";
RL   Mol. Cell. Endocrinol. 118:85-94(1996).
RN   [4]
RP   SYNTHESIS.
RX   MEDLINE=91250367; PubMed=2040595;
RA   Buellesbach E.E., Schwabe C.;
RT   "Total synthesis of human relaxin and human relaxin derivatives by
RT   solid-phase peptide synthesis and site-directed chain combination.";
RL   J. Biol. Chem. 266:10754-10761(1991).
RN   [5]
RP   PARTIAL SEQUENCE OF 25-51.
RC   TISSUE=Semen;
RX   MEDLINE=92241162; PubMed=1572287;
RA   Winslow J.W., Shih A., Bourell J.H., Weiss G., Reed B., Stults J.T.,
RA   Goldsmith L.T.;
RT   "Human seminal relaxin is a product of the same gene as human luteal
RT   relaxin.";
RL   Endocrinology 130:2660-2668(1992).
RN   [6]
RP   SEQUENCE OF 25-53 AND 162-185.
RX   MEDLINE=91167739; PubMed=2076464;
RA   Winslow J.W., Griffin P.R., Rinderknecht E., Vandlen R.L.;
RT   "Structural characterization by mass spectrometry of native and
RT   recombinant human relaxin.";
RL   Biomed. Environ. Mass Spectrom. 19:655-664(1990).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX   MEDLINE=92015205; PubMed=1656049;
RA   Eigenbrot C., Randal M., Quan C., Burnier J., O'Connell L.,
RA   Rinderknecht E., Kossiakoff A.A.;
RT   "X-ray structure of human relaxin at 1.5 A. Comparison to insulin and
RT   implications for receptor binding determinants.";
RL   J. Mol. Biol. 221:15-21(1991).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC   PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC   INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC   PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC   DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC   produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in the ovary during pregnancy. Also
CC   expressed in placenta, decidua and prostate.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X00948; CAA25460.1; -.
DR   EMBL; AL135786; CAC04177.1; -.
DR   EMBL; AL135786; CAC04176.1; -.
DR   EMBL; S83200; AAD14429.1; -.
DR   EMBL; A17315; CAA01324.1; -.
DR   EMBL; A06925; CAA00602.1; -.
DR   PIR; A60982; A60982.
DR   PDB; 6RLX; 31-OCT-93.
DR   Genew; HGNC:10027; RLN2.
DR   MIM; 179740; -.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Multigene family; Signal; 3D-structure;
KW   Alternative splicing.
FT   SIGNAL   1   24
FT   CHAIN   25   53   RELAXIN B CHAIN.
FT   PROPEP   56   157   CONNECTING PEPTIDE.
FT   CHAIN   162   185   RELAXIN A CHAIN.
FT   MOD_RES   162   162   PYRROLIDONE CARBOXYLIC ACID.
FT   DISULFID   35   172   INTERCHAIN.
FT   DISULFID   47   185   INTERCHAIN.
FT   DISULFID   171   176
FT   VARSPLIC   71   117   EIVPSFINKDTETINMMSEFVANLPQELKLTLSEMQPALPQ
FT   LQQHVP -> GDFIQTVSLGISPDGGKALRTGSCFTREFLG
FT   ALSKLCHPSSTKIQKP (IN ISOFORM 2).
FT   VARSPLIC   118   185   MISSING (IN ISOFORM 2).
FT   HELIX   27   29
FT   STRAND   31   32
FT   HELIX   36   49
```
```

```
FT   HELIX       163    173
FT   TURN        174    174
FT   STRAND      176    177
FT   HELIX       178    182
FT   TURN        183    184
SQ   SEQUENCE   185 AA;  21042 MW;  AC73DBDE2090091B CRC64;

Alignment Scores:
Pred. No.:                0.00116        Length:         185
Score:                    105.50         Matches:         43
Percent Similarity:       35.67%         Conservative:    13
Best Local Similarity:    27.39%         Mismatches:      44
Query Match:              16.61%         Indels:          57
DB:                       1              Gaps:             6

US-09-518-842-1_COPY_76_417 (1-342) x REL2_HUMAN (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         42 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::               ||||||                          ||||||
Db         62 AlaProGln--------------ThrPro----------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::      |||||||||   :::   :::         |||
Db         69 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              ||||||:::   ||||||      |||||||||      |||||||||   ||||||:::
Db         89 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 108

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db        109 LeuProGlnLeuGlnGlnHisValProValLeuLysAspSerSerLeuLeuPheGluGlu 128

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                 ||||||:::|||
Db        129 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 148

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||||||||||||
Db        149 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                  ||||||   |||   |||         :::      :::   |||
Db        169 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 185

RESULT 8
REL1_HUMAN
ID   REL1_HUMAN     STANDARD;      PRT;   185 AA.
AC   P04808; Q99936;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Prorelaxin H1 precursor.
GN   RLN1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=85051298; PubMed=6548702;
RA   Hudson P., John M., Crawford R., Haralambidis J., Scanlon D.,
RA   Gorman J., Tregear G., Shine J., Niall H.;
RT   "Relaxin gene expression in human ovaries and the predicted structure
RT   of a human preprorelaxin by analysis of cDNA clones.";
RL   EMBO J. 3:2333-2339(1984).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=83141755; PubMed=6298628;
RA   Hudson P., Haley J., John M., Cronk M., Crawford R., Haralambidis J.,
RA   Tregear G., Shine J., Niall H.;
RT   "Structure of a genomic clone encoding biologically active human
RT   relaxin.";
RL   Nature 301:628-631(1983).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Sehra H.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Prostate;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE
RP   SPECIFICITY.
RC   TISSUE=Prostate;
RX   MEDLINE=96328899; PubMed=8735594;
RA   Gunnersen J.M., Fu P., Roche P.J., Tregear G.W.;
RT   "Expression of human relaxin genes: characterization of a novel
RT   alternatively-spliced human relaxin mRNA species.";
RL   Mol. Cell. Endocrinol. 118:85-94(1996).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Prostate. Not expressed in placenta, decidua
CC       or ovary.
CC   -!- MISCELLANEOUS: H1 RELAXIN MAY BE A PSEUDOGENE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00949; CAA25461.1; -.
DR   EMBL; V00578; CAA23839.1; -.
DR   EMBL; V00577; CAA23838.1; -.
DR   EMBL; A06926; CAA00603.1; -.
DR   EMBL; A06846; CAA00599.1; -.
DR   EMBL; A07364; CAA00658.1; -.
DR   EMBL; A17329; CAA01325.1; -.
DR   EMBL; AL135786; CAC04179.1; -.
DR   EMBL; BC005956; AAH05956.1; -.
DR   EMBL; S83200; AAD14429.1; -.
DR   PIR; A44559; A44559.
DR   HSSP; P04090; 6RLX.
DR   Genew; HGNC:10026; RLN1.
DR   MIM; 179730; -.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   PRINTS; PR00276; INSULINA.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Multigene family; Signal;
KW   Alternative splicing; Polymorphism.
FT   SIGNAL        1     22       PROBABLE.
FT   CHAIN        23     53       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       56    158       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       163    185       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     35    172       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     47    185       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    171    176       BY SIMILARITY.
```

```
FT   VARSPLIC     71    117       EIVPSFINKDTETIIIMLEFIANLPPELKAALSERQPSLPE
FT                                LQQYVP -> GDFIQTVSLGISPDGGKALRTGSCFTREFLG
FT                                ALSKLYHPSSTKIQKL (IN ISOFORM 2).
FT   VARSPLIC    118    185       MISSING (IN ISOFORM 2).
FT   VARIANT      28     28        K -> M (IN DBSNP:618066).
FT                                /FTId=VAR_011962.
SQ   SEQUENCE   185 AA;  21145 MW;  B318628ABFEC7142 CRC64;

Alignment Scores:
Pred. No.:                0.00193        Length:          185
Score:                    103.50         Matches:         44
Percent Similarity:       35.67%         Conservative:    12
Best Local Similarity:    28.03%         Mismatches:      44
Query Match:              16.30%         Indels:          57
DB:                       1              Gaps:            6

US-09-518-842-1_COPY_76_417 (1-342) x REL1_HUMAN (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||        :::   |||      ||||||
Db         42 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                |||:::              ||||||                          ||||||
Db         62 AlaProGln-------------ThrPro----------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::    ||||||||||   :::   :::
Db         69 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleIleMetLeu 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||||   ||||||   ||||||||||||      |||||||||   |||||||||
Db         89 GluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSerGluArgGlnProSer 108

Qy        235 TTG------------------------------------------------------- 237
              |||
Db        109 LeuProGluLeuGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGlu 128

Qy        238 ---AAGAAAATAATACTTTCCCGC---------------------------------- 258
              ||||||:::|||   :::|||
Db        129 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeu 148

Qy        259 ----------------------------AAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                           |||||||||
Db        149 LysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPhe 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                |||||   :::   |||         :::      |||   |||
Db        169 GluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaLysTyrCys 185

RESULT 9
RELX_MACMU
ID   RELX_MACMU     STANDARD;      PRT;   185 AA.
AC   P19884;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor.
GN   RLN.
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9544;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90073957; PubMed=2590381;
RA   Crawford R.J., Hammond V.E., Roche P.J., Johnston P.D., Tregear G.W.;
RT   "Structure of rhesus monkey relaxin predicted by analysis of the
RT   single-copy rhesus monkey relaxin gene.";
RL   J. Mol. Endocrinol. 3:169-174(1989).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR   PIR; A34936; A34936.
DR   HSSP; P04090; 6RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23     53       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP      56    157       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN      161    185       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID    35    172       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    47    185       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   171    176       BY SIMILARITY.
SQ   SEQUENCE   185 AA;  20895 MW;  7E3C5D21B57E185C CRC64;

Alignment Scores:
Pred. No.:                0.00193        Length:          185
Score:                    103.50         Matches:         41
Percent Similarity:       34.39%         Conservative:    13
Best Local Similarity:    26.11%         Mismatches:      46
Query Match:              16.30%         Indels:          57
DB:                       1              Gaps:            5

US-09-518-842-1_COPY_76_417 (1-342) x RELX_MACMU (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||:::::      ||||||        ||||||      |||:::
Db         42 AlaGlnIleAlaIleCysGlyLysSerThrLeuGlyLysArgSerLeuAsnGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                |||   |||                                              ||||||
Db         62 AlaProLeuLysPro-----------------------------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::      |||:::|||   :::   :::         |||
Db         69 AlaAlaGluIleValProSerLeuIleAsnGlnAspThrGluThrIleAsnMetMetSer 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||:::   ||||||      |||||||||      |||||||||   ||||||:::
Db         89 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluArgGlnProAla 108

Qy        235 TTG------------------------------------------------------- 237
              |||
Db        109 LeuSerGluLeuGlnGlnHisValProValLeuLysAspSerAsnLeuSerPheGluGlu 128

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
              ||||||||||||
Db        129 PheLysLysIleIleArgLysArgGlnSerGluAlaThrAspSerSerProSerGluLeu 148

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||:::||||||
Db        149 ArgSerLeuGlyLeuAspThrHisSerArgArgLysArgGlnLeuTyrMetThrLeuSer 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                |||||   :::   |||         :::      |||   |||
Db        169 AsnLysCysCysHisIleGlyCysThrLysLysSerLeuAlaLysPheCys 185

RESULT 10
REL1_PANTR
ID   REL1_PANTR     STANDARD;      PRT;   166 AA.
AC   P51454;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin H1 precursor (Fragment).
GN   RNL1 OR RLX1.
```

```
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94238260; PubMed=8182365;
RA   Evans B.A., Fu P., Tregear G.W.;
RT   "Characterization of two relaxin genes in the chimpanzee.";
RL   J. Endocrinol. 140:385-392(1994).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. MAY BE
CC       INVOLVED IN REMODELING OF CONNECTIVE TISSUES DURING PREGNANCY,
CC       PROMOTING GROWTH OF PUBIC LIGAMENTS AND RIPENING OF THE CERVIX.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE CORPUS LUTEUM OF PREGNANCY
CC       BUT NOT IN THE PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z27225; CAA81739.1; -.
DR   HSSP; P04090; 6RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Multigene family; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1      5       POTENTIAL.
FT   CHAIN         6     34       RELAXIN B CHAIN (PROBABLE).
FT   PROPEP       37    139       CONNECTING PEPTIDE (PROBABLE).
FT   CHAIN       143    166       RELAXIN A CHAIN (PROBABLE).
FT   DISULFID     16    153       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     28    166       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    152    157       BY SIMILARITY.
SQ   SEQUENCE   166 AA;  18730 MW;  7F469B1FB9259F4F CRC64;

Alignment Scores:
Pred. No.:              0.00407      Length:          166
Score:                  100.50       Matches:         41
Percent Similarity:     36.94%       Conservative:    17
Best Local Similarity:  26.11%       Mismatches:      42
Query Match:            15.83%       Indels:          57
DB:                     1            Gaps:            7

US-09-518-842-1_COPY_76_417 (1-342) x REL1_PANTR (1-166)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         23 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 42

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::               ||||||                        ||||||
Db         43 AlaProGln--------------ThrPro-----------------------ArgPro 49

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::     |||||||||   :::   :::
Db         50 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeu 69

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||||   ||||||   ||||||||||||      |||||||||   |||||||||
Db         70 GluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSerGluArgGlnProSer 89

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db         90 LeuProGluProGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGlu 109

Qy        238 ---AAGAAAATAATACTTTCCCGCAAAAAGAGAAGT---------------------- 270
                 ||||||:::|||   :::|||:::      :::
Db        110 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeu 129

Qy        271 ------------------------GGACGTCACAGATTTGATCCATTC------------ 294
                                        :::   :::      |||:::
Db        130 LysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgGlnProTyrValAlaLeuPhe 149

Qy        295 ------TGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                    ||||||   :::   |||         :::           |||
Db        150 GluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaAsnTyrCys 166

RESULT 11
RELX_RAT
ID   RELX_RAT       STANDARD;      PRT;   186 AA.
AC   P01347;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor.
GN   RLN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81197624; PubMed=7231533;
RA   Hudson P., Haley J., Cronk M., Shine J., Niall H.;
RT   "Molecular cloning and characterization of cDNA sequences coding for
RT   rat relaxin.";
RL   Nature 291:127-131(1981).
RN   [2]
RP   SEQUENCE OF 23-57 AND 163-188.
RX   MEDLINE=81090283; PubMed=7004862;
RA   John M.J., Borjesson B.W., Walsh J.R., Niall H.D.;
RT   "Limited sequence homology between porcine and rat relaxins:
RT   implications for physiological studies.";
RL   Endocrinology 108:726-729(1981).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN
CC       TO PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J00780; AAA42029.1; -.
DR   EMBL; V01264; CAA24578.1; -.
DR   PIR; A01614; RXRT.
DR   HSSP; P01348; 1RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23     57       RELAXIN B CHAIN.
FT   PROPEP       58    158       CONNECTING PEPTIDE.
FT   CHAIN       163    186       RELAXIN A CHAIN.
FT   MOD_RES     163    163       PYRROLIDONE CARBOXYLIC ACID.
FT   DISULFID     36    173       INTERCHAIN (BY SIMILARITY).
```

```
FT   DISULFID     48    186       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    172    177       BY SIMILARITY.
SQ   SEQUENCE   186 AA;  20489 MW;  08BAEC79BCF0E80F CRC64;

Alignment Scores:
Pred. No.:                0.00888        Length:        186
Score:                    97.50          Matches:       38
Percent Similarity:       31.79%         Conservative:  10
Best Local Similarity:    25.17%         Mismatches:    48
Query Match:              15.35%         Indels:        55
DB:                       1              Gaps:          4

US-09-518-842-1_COPY_76_417 (1-342) x RELX_RAT (1-186)

Qy         16 TGTGGTCCCCGATTTGGAAAACACTTGCTGTCATAT---TGCCCCATGCCTGAGAAGACA 72
              ||||||    |||:::    ||||||       |||   |||
Db         48 CysGlyAlaSerValGlyArgLeuAlaLeuSerGlnGluGluProAlaPro--------- 64

Qy         73 TTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCA 132
                                        |||                 |||:::|||
Db         65 -------------------------LeuAlaArgGlnAlaThrAlaGluValValPro 75

Qy        133 ACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTG 192
              :::      |||||||||   :::           |||   :::   :::|||||||||
Db         76 SerPheIleAsnLysAspAlaGluProPheAspMetThrLeuLysCysLeuProAsnLeu 95

Qy        193 TCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAG---------------------- 228
              |||   |||   |||      ||||||||||||:::
Db         96 SerGluGluArgLysAlaAlaLeuSerGluGlyArgAlaProPheProGluLeuGlnGln 115

Qy        228 ------------------------------------------------------------ 228

Db        116 HisAlaProAlaLeuSerAspSerValValSerLeuGluGlyPheLysLysThrPheHis 135

Qy        229 --------------------------------CCATCATTGAAGAAAATA--------- 246
                                              |||   ||||||   :::
Db        136 AsnGlnLeuGlyGluAlaGluAspGlyGlyProProGluLeuLysTyrLeuGlySerAsp 155

Qy        247 ATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTA 306
              ||||||||||||||||                                ||||||   :::
Db        156 AlaGlnSerArgLysLysArgGlnSerGlyAlaLeuLeuSerGluGlnCysCysHisIle 175

Qy        307 ATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
              |||       :::      ||||||||
Db        176 GlyCysThrArgArgSerIleAlaLysLeuCys 186

RESULT 12
RELX_CAVPO
ID   RELX_CAVPO     STANDARD;      PRT;   160 AA.
AC   P51453;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prorelaxin precursor (Fragment).
GN   RLN.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Endometrium;
RX   MEDLINE=92164503; PubMed=1537282;
RA   Lee Y.A., Bryant-Greenwood G.D., Mandel M., Greenwood F.C.;
RT   "The complementary deoxyribonucleic acid sequence of guinea pig
RT   endometrial prorelaxin.";
RL   Endocrinology 130:1165-1172(1992).
CC   -!- FUNCTION: RELAXIN IS AN OVARIAN HORMONE THAT ACTS WITH ESTROGEN TO
CC       PRODUCE DILATATION OF THE BIRTH CANAL IN MANY MAMMALS. IT BEARS
CC       MATURE YOUNG, AND ALLOWS SEPARATION OF THE PELVIC BONES.
CC   -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC       DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE ENDOMETRIUM DURING PREGNANCY
CC       AND IN MAMMARY GLAND DURING LACTATION.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S85964; AAB21586.1; -.
DR   HSSP; P01348; 1RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
KW   Insulin family; Hormone.
FT   NON_TER       1      1
FT   CHAIN        <1     33       RELAXIN B CHAIN (POTENTIAL).
FT   PROPEP       34    133       CONNECTING PEPTIDE (POTENTIAL).
FT   CHAIN       137    160       RELAXIN A CHAIN (POTENTIAL).
FT   DISULFID     10    147       INTERCHAIN (BY SIMILARITY).
FT   DISULFID     22    160       INTERCHAIN (BY SIMILARITY).
FT   DISULFID    146    151       BY SIMILARITY.
SQ   SEQUENCE   160 AA;  17892 MW;  E58DF84036467028 CRC64;

Alignment Scores:
Pred. No.:                1.79           Length:        160
Score:                    76.50          Matches:       31
Percent Similarity:       35.09%         Conservative:  9
Best Local Similarity:    27.19%         Mismatches:    31
Query Match:              12.05%         Indels:        43
DB:                       1              Gaps:          3

US-09-518-842-1_COPY_76_417 (1-342) x RELX_CAVPO (1-160)

Qy         97 CTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAA 156
              :::|||   ||||||:::      |||::::::   :::|||   ||||||:::
Db         39 IleLeuGlySerGlyGlnSerAlaGluIleMetProSerSerIleAsnLysGluValAsp 58

Qy        157 GCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTG 216
              :::|||            |||   |||   ||||||      ||||||:::      |||
Db         59 SerLeuAsnMetLeu---GluSerIleAlaAsnLeuProGluGluLeuArgAlaMetLeu 77

Qy        217 TCTGAAGGGCAGCCATCA------------------------------------------ 234
                |||   |||||||||
Db         78 ProGluLysGlnProSerSerProGlnLeuGlnGlnTyrValProAlaLeuLysAsnSer 97

Qy        235 ------------------TTGAAGAAAATAATACTTTCCCGC------------------ 258
                               |||   |||||||||      |||
Db         98 AsnValAlaValLysGluLeuAsnLysIleIleArgGlyArgGlnGluGluAlaGluAsp 117

Qy        259 ------------------------------------------------AAAAAGAGAAGT 270
                                                             |||||||||
Db        118 AsnSerHisSerLeuLeuLysAspPheAsnLeuAsnIleTyrSerProLysLysArgGln 137

Qy        271 GGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGT 312
                                     ||||||:::|||   |||
Db        138 LeuAspMetThrValSerGluLysCysCysGlnValGlyCys 151

RESULT 13
DPOL_HPBDU
ID   DPOL_HPBDU     STANDARD;      PRT;   836 AA.
AC   P03162;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   P protein [Includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-
DE   directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].
GN   P.
OS   Duck hepatitis B virus (DHBV).
```

```
OC   Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX   NCBI_TaxID=12639;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84138772; PubMed=6699938;
RA   Mandart E., Kay A., Galibert F.;
RT   "Nucleotide sequence of a cloned duck hepatitis B virus genome:
RT   comparison with woodchuck and human hepatitis B virus sequences.";
RL   J. Virol. 49:782-792(1984).
RN   [2]
RP   SEQUENCE OF 794-836 FROM N.A.
RX   MEDLINE=84216498; PubMed=6328037;
RA   Molnar-Kimber K.L., Summers J.W., Mason W.S.;
RT   "Mapping of the cohesive overlap of duck hepatitis B virus DNA and of
RT   the site of initiation of reverse transcription.";
RL   J. Virol. 51:181-191(1984).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC       phosphomonoester.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K01834; AAA45742.1; -.
DR   PIR; A00710; JDVLD.
DR   InterPro; IPR001462; DNApol_viral_C.
DR   InterPro; IPR000201; DNApol_viral_N.
DR   InterPro; IPR000477; RVTse.
DR   Pfam; PF00078; rvt; 1.
DR   Pfam; PF00242; DNA_pol_viral_N; 1.
DR   Pfam; PF00336; DNA_pol_viral_C; 1.
DR   ProDom; PD000814; DNApol_viral_C; 1.
KW   Transferase; RNA-directed DNA polymerase; DNA-directed DNA polymerase;
KW   Hydrolase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SQ   SEQUENCE   836 AA;  95275 MW;  FFEB57CFF549A4F4 CRC64;

Alignment Scores:
Pred. No.:                3.46          Length:        836
Score:                    75.00         Matches:       25
Percent Similarity:       42.42%        Conservative:  17
Best Local Similarity:    25.25%        Mismatches:    27
Query Match:              11.81%        Indels:        30
DB:                       1             Gaps:          6

US-09-518-842-1_COPY_76_417 (1-342) x DPOL_HPBDU (1-836)

Qy         14 GATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGA------ 67
              |||:::|||      :::   ||||||          :::   ||||||
Db        295 AspMetValArgGlnValSerAsnThrArgSerArgValArgProCysAlaAsnAsnGly 314

Qy         68 ------AGACATTCACCACCACCCCAGGAGGGT------GGCTGCTGGAATCTGGACGTC 115
                  :::|||      ||||||:::   |||         ||||||     |||:::
Db        315 GlyAspLysHis------ProProGluSerGlySerLeuAlaCysTrp--GlyGlyLysG 332

Qy        116 CCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAG 175
                  ::::::   :::   :::::::::|||      |||
Db        332 luSerArgIleIleLysSerAspSerSerArgAspSerSerAla---------------- 346

Qy        176 AATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTG---TCTGAAGGGCAGCCAT 232
                                          |||:::   |||   |||:::|||
Db        347 ---------------------------------ProValAspSerArgGlyArgProL 355

Qy        233 CATTGAAGAAAATAATA------CTTTCCCGCAAAAAGAGAAGTGGACGTCAC 279
                  :::          ||||||||||:::|||   :::|||   |||
Db        355 ysSerThrArgSerPheSerProLeuSerArgArgLysThrThrGlyAsnHis 372

RESULT 14
PMPD_CHLTR
ID   PMPD_CHLTR     STANDARD;      PRT;  1531 AA.
AC   O84818;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable outer membrane protein pmpD precursor (Polymorphic membrane
DE   protein D).
GN   PMPD OR CT812.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/Cx;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001353; AAC68408.1; -.
DR   PHCI-2DPAGE; O84818; -.
DR   InterPro; IPR003368; Chlamydia_PMP.
DR   InterPro; IPR003357; OMP.
DR   Pfam; PF02385; OMP; 1.
DR   Pfam; PF02415; DUF145; 1.
KW   Outer membrane; Signal; Multigene family; Complete proteome.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1531       PROBABLE OUTER MEMBRANE PROTEIN PMPD.
SQ   SEQUENCE   1531 AA;  160747 MW;  584ABD11E241EF22 CRC64;

Alignment Scores:
Pred. No.:                4.35          Length:        1531
Score:                    74.50         Matches:       29
Percent Similarity:       44.79%        Conservative:  14
Best Local Similarity:    30.21%        Mismatches:    26
Query Match:              11.73%        Indels:        27
DB:                       1             Gaps:          5

US-09-518-842-1_COPY_76_417 (1-342) x PMPD_CHLTR (1-1531)

Qy          4 GAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTG---------------------- 42
              |||   :::|||   |||      ||||||:::::::::
Db        472 GluTyrGlnGlyGlyGlyAlaLeuPheGlyGluAsnIleSerLeuSerGluAsnAlaGly 491

Qy         43 ---CTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACC------------CCA 87
                 |||::::::               |||||||||   :::
Db        492 ValLeuThrPheLysAspAsnIleValLysThrPheAlaSerAsnGlyLysIleLeuGly 511

Qy         88 GGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 147
              ||||||   :::|||   :::|||:::              |||      :::||||||
Db        512 GlyGlyAlaIleLeuAlaThrGlyLys------------ValGluIleThrAsnAsnSer 527

Qy        148 GATGGA-----------------------------CAAGCCTTAGGTACGACATCAGAA 177
              :::|||                            |||||||||   |||      |||
Db        528 GluGlyIleSerPheThrGlyAsnAlaArgAlaProGlnAlaLeuProThrGlnGluGlu 547

Qy        178 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 225
              |||   |||      |||   :::      :::||||||||||   |||
```

```
Db        548 Phe---ProLeuPheSerLysLysGluGlyArgProLeuSerSerGly 562

RESULT 15
Y188_HUMAN
ID   Y188_HUMAN     STANDARD;      PRT;   899 AA.
AC   Q14693;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein KIAA0188 (Fragment).
GN   KIAA0188.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
CC   -!- SIMILARITY: TO HUMAN KIAA0249 AND YEAST SMP2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D80010; BAA11505.1; -.
DR   Genew; HGNC:13345; LPIN1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   899 AA;  99756 MW;  AF81E5B095035924 CRC64;

Alignment Scores:
Pred. No.:              5.13          Length:          899
Score:                  73.50         Matches:         20
Percent Similarity:     47.06%        Conservative:    12
Best Local Similarity:  29.41%        Mismatches:      35
Query Match:            11.57%        Indels:          1
DB:                     1             Gaps:            1

US-09-518-842-1_COPY_76_417 (1-342) x Y188_HUMAN (1-899)

Qy         79 ACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCC 138
              |||     ||||||   ||||||:::      :::|||:::      :::
Db        333 ThrLeuValGlyGlyAlaLeuLeuAspGlnAsnLysProGlnThrGluMetGlnPheVal 352

Qy        139 AACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCA 198
              |||   :::|||   :::   ||||||      :::      :::|||   :::
Db        353 AsnGluGluAspLeuGluThrLeuGlyAlaAlaAlaProLeuLeuProMetIleGlu--- 371

Qy        199 GAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGC 258
              |||||||||   |||   :::            :::      |||
Db        372 GluLeuLysProProSerAlaSerValValGlnThrAlaAsnLysThrAspSerProSer 391

Qy        259 AAAAAGAGAAGTGGACGTCACAGA 282
              :::||||||      |||   |||
Db        392 ArgLysArgAspLysArgSerArg 399


Search completed: June 14, 2003, 19:43:32
Job time : 14 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:43:16 ; Search time 12.5 Seconds
(without alignments)
1610.022 Million cell updates/sec

Title: US-09-518-842-1_COPY_76_417
Perfect score: 635
Sequence: 1 GCAGAGCTGAGGGGATGTGG.....CTTCAGTTAAATTATGTACA 342

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09518842/runat_14062003_175812_12308/app_query.fasta_1.519
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09518842_@CGN_1_1_28_@runat_14062003_175812_12308 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	619	97.5	124	3	US-08-991-890-5	Sequence 5, Appli
2	619	97.5	139	3	US-08-950-720A-15	Sequence 15, Appl
3	619	97.5	139	3	US-08-991-890-2	Sequence 2, Appli
4	619	97.5	159	3	US-08-991-890-4	Sequence 4, Appli
5	613	96.5	139	4	US-09-599-564A-2	Sequence 2, Appli
6	608	95.7	174	4	US-09-174-465D-2	Sequence 2, Appli
7	255	40.2	51	4	US-09-174-465D-10	Sequence 10, Appl
8	255	40.2	51	4	US-09-599-564A-10	Sequence 10, Appl
9	193	30.4	41	4	US-09-174-465D-8	Sequence 8, Appli
10	193	30.4	41	4	US-09-599-564A-8	Sequence 8, Appli
11	193	30.4	41	4	US-09-201-227A-27	Sequence 27, Appl
12	166	26.1	30	4	US-09-174-465D-12	Sequence 12, Appl
13	166	26.1	30	4	US-09-599-564A-12	Sequence 12, Appl
14	147	23.1	25	4	US-09-201-227A-28	Sequence 28, Appl
15	105.5	16.6	150	1	US-08-443-568B-12	Sequence 12, Appl
16	105.5	16.6	150	5	PCT-US94-06997-12	Sequence 12, Appl
17	105.5	16.6	162	6	5464756-18	Patent No. 5464756
18	105.5	16.6	185	3	US-08-950-720A-12	Sequence 12, Appl
19	103.5	16.3	185	3	US-08-950-720A-13	Sequence 13, Appl
20	99.5	15.7	164	6	5464756-20	Patent No. 5464756
c 21	73	12.2	994	3	US-08-699-103B-3	Sequence 3, Appli
c 22	73	12.2	994	4	US-09-229-059-3	Sequence 3, Appli
23	71.5	11.3	778	4	US-09-556-877-193	Sequence 193, App
24	71.5	11.3	778	4	US-09-620-412C-193	Sequence 193, App
25	71.5	11.3	1530	4	US-09-556-877-178	Sequence 178, App
26	71.5	11.3	1530	4	US-09-620-412C-178	Sequence 178, App
27	69	10.9	2409	6	5180808-2	Patent No. 5180808
28	67	10.6	313	1	US-08-078-683A-4	Sequence 4, Appli
29	66.5	10.5	846	4	US-09-134-001C-3067	Sequence 3067, Ap
c 30	66	11.0	433	4	US-09-046-158A-2	Sequence 2, Appli
c 31	66	11.0	897	1	US-07-960-389-2	Sequence 2, Appli
32	65.5	10.3	1183	4	US-09-134-001C-3530	Sequence 3530, Ap
33	64.5	10.2	156	4	US-09-134-001C-5102	Sequence 5102, Ap
34	64.5	10.2	704	1	US-08-188-582-18	Sequence 18, Appl
35	64.5	10.2	704	1	US-08-646-715-18	Sequence 18, Appl
c 36	64	10.7	188	3	US-09-332-934-2	Sequence 2, Appli
37	63.5	10.0	383	4	US-09-134-001C-3068	Sequence 3068, Ap
38	63.5	10.0	572	2	US-08-669-524-8	Sequence 8, Appli
39	63.5	10.0	619	4	US-09-423-890-4	Sequence 4, Appli
40	63	9.9	446	2	US-08-244-205-5	Sequence 5, Appli
41	63	9.9	446	4	US-09-161-994A-5	Sequence 5, Appli
42	63	9.9	446	5	PCT-US92-10284-5	Sequence 5, Appli
43	63	9.9	446	5	PCT-US94-01321-10	Sequence 10, Appl
44	63	9.9	535	4	US-09-269-731-4	Sequence 4, Appli
45	62.5	9.8	554	2	US-08-524-051-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-991-890-5
; Sequence 5, Application US/08991890
; Patent No. 6114307
;  GENERAL INFORMATION:
;    APPLICANT:  Jaspers, Stephen R.
;    APPLICANT:  Sprugel, Katherine H.
;    APPLICANT:  Ren, Hong Ping
;    APPLICANT:  Humes, Jacqueline M.
;    APPLICANT:  Hoffman, Ross C.
;    APPLICANT:  Conklin, Darrell C.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  STIMULATING PANCREATIC ISLET CELL REGENERATION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,890
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/033,003
;      FILING DATE:  December 16, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sawislak, Deborah A
;      REGISTRATION NUMBER:  37,438
```

```
;      REFERENCE/DOCKET NUMBER:  96-41
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6672
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  124 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-991-890-5

Alignment Scores:
Pred. No.:              3.43e-69        Length:         124
Score:                  619.00          Matches:        114
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            97.48%          Indels:         0
DB:                     3               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-08-991-890-5 (1-124)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 30

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 50

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 70

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 90

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 110

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||
Db        111 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 124

RESULT 2
US-08-950-720A-15
; Sequence 15, Application US/08950720A
; Patent No. 6046028
;  GENERAL INFORMATION:
;    APPLICANT:  Conklin, Darrell C.
;    APPLICANT:  Lofton-Day, Catherine E.
;    APPLICANT:  Lok, Si
;    APPLICANT:  Jaspers, Stephen R.
;    TITLE OF INVENTION:  INSULIN HOMOLOG
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/950,720A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sawislak, Deborah A
;      REGISTRATION NUMBER:  37,438
;      REFERENCE/DOCKET NUMBER:  96-09
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6672
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  139 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6046028e
US-08-950-720A-15

Alignment Scores:
Pred. No.:              3.59e-69        Length:         139
Score:                  619.00          Matches:        114
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            97.48%          Indels:         0
DB:                     3               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-08-950-720A-15 (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 3
US-08-991-890-2
; Sequence 2, Application US/08991890
; Patent No. 6114307
;  GENERAL INFORMATION:
;    APPLICANT:  Jaspers, Stephen R.
;    APPLICANT:  Sprugel, Katherine H.
;    APPLICANT:  Ren, Hong Ping
;    APPLICANT:  Humes, Jacqueline M.
;    APPLICANT:  Hoffman, Ross C.
;    APPLICANT:  Conklin, Darrell C.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  STIMULATING PANCREATIC ISLET CELL REGENERATION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
```

```
;       STATE:  WA
;       COUNTRY:  USA
;       ZIP:  98102
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/991,890
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/033,003
;       FILING DATE:  December 16, 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sawislak, Deborah A
;       REGISTRATION NUMBER:  37,438
;       REFERENCE/DOCKET NUMBER:  96-41
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-442-6672
;       TELEFAX:  206-442-6678
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  139 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:  internal
US-08-991-890-2

Alignment Scores:
Pred. No.:                3.59e-69        Length:         139
Score:                    619.00          Matches:        114
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              97.48%          Indels:         0
DB:                       3               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-08-991-890-2 (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 4
US-08-991-890-4
; Sequence 4, Application US/08991890
; Patent No. 6114307
;  GENERAL INFORMATION:
;    APPLICANT:  Jaspers, Stephen R.
;    APPLICANT:  Sprugel, Katherine H.
;    APPLICANT:  Ren, Hong Ping
;    APPLICANT:  Humes, Jacqueline M.
;    APPLICANT:  Hoffman, Ross C.
;    APPLICANT:  Conklin, Darrell C.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  STIMULATING PANCREATIC ISLET CELL REGENERATION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,890
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/033,003
;      FILING DATE:  December 16, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sawislak, Deborah A
;      REGISTRATION NUMBER:  37,438
;      REFERENCE/DOCKET NUMBER:  96-41
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6672
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  159 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-991-890-4

Alignment Scores:
Pred. No.:                3.79e-69        Length:         159
Score:                    619.00          Matches:        114
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              97.48%          Indels:         0
DB:                       3               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-08-991-890-4 (1-159)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 65

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 85

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 105

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 125

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        126 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 145

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||||
Db        146 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 159

RESULT 5
US-09-599-564A-2
; Sequence 2, Application US/09599564A
; Patent No. 6362318
; GENERAL INFORMATION:
;  APPLICANT: KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
;  FILE REFERENCE: 017753-127
;  CURRENT APPLICATION NUMBER: US/09/599,564A
;  CURRENT FILING DATE:  2000-06-23
;  PRIOR APPLICATION NUMBER: 09/174,465
;  PRIOR FILING DATE: 1998-10-19
;  PRIOR APPLICATION NUMBER: US 08/482,842
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 139
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-2

Alignment Scores:
Pred. No.:                2.02e-68        Length:         139
Score:                    613.00          Matches:        113
Percent Similarity:       99.12%          Conservative:   0
Best Local Similarity:    99.12%          Mismatches:     1
Query Match:              96.54%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-599-564A-2 (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerLysAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 6
US-09-174-465D-2
; Sequence 2, Application US/09174465D
; Patent No. 6180364
; GENERAL INFORMATION:
;  APPLICANT:  KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION:  NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
;  FILE REFERENCE:  017753-103
;  CURRENT APPLICATION NUMBER: US/09/174,465D
;  CURRENT FILING DATE:  1998-10-19
;  PRIOR APPLICATION NUMBER:  US 08/482,842
;  PRIOR FILING DATE:  1995-06-07
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 174
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-2

Alignment Scores:
Pred. No.:                9.33e-68        Length:         174
Score:                    608.00          Matches:        113
Percent Similarity:       99.12%          Conservative:   0
Best Local Similarity:    99.12%          Mismatches:     1
Query Match:              95.75%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-174-465D-2 (1-174)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 80

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 100

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        101 GluMetValSerThrSerLysAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 120

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 140

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 160

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||||
Db        161 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 174

RESULT 7
US-09-174-465D-10
; Sequence 10, Application US/09174465D
; Patent No. 6180364
; GENERAL INFORMATION:
;  APPLICANT:  KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION:  NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
;  FILE REFERENCE:  017753-103
;  CURRENT APPLICATION NUMBER: US/09/174,465D
```

```
;  CURRENT FILING DATE:  1998-10-19
;  PRIOR APPLICATION NUMBER:  US 08/482,842
;  PRIOR FILING DATE:  1995-06-07
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 51
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-10

Alignment Scores:
Pred. No.:                8.36e-24        Length:         51
Score:                    255.00          Matches:        51
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.16%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-174-465D-10 (1-51)

Qy         100 CTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCC 159
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 LeuGluSerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAla 20

Qy         160 TTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 219
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 LeuGlyThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSer 40

Qy         220 GAAGGGCAGCCATCATTGAAGAAAATAATACTT 252
               ||||||||||||||||||||||||||||||||
Db          41 GluGlyGlnProSerLeuLysLysIleIleLeu 51

RESULT 8
US-09-599-564A-10
; Sequence 10, Application US/09599564A
; Patent No: 6362318
; GENERAL INFORMATION:
;  APPLICANT: KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
;  FILE REFERENCE: 017753-127
;  CURRENT APPLICATION NUMBER: US/09/599,564A
;  CURRENT FILING DATE:  2000-06-23
;  PRIOR APPLICATION NUMBER: 09/174,465
;  PRIOR FILING DATE: 1998-10-19
;  PRIOR APPLICATION NUMBER: US 08/482,842
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 51
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-10

Alignment Scores:
Pred. No.:                8.36e-24        Length:         51
Score:                    255.00          Matches:        51
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              40.16%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-599-564A-10 (1-51)

Qy         100 CTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCC 159
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 LeuGluSerGlyArgProLysGluMetValSerThrSerAsnAsnLysAspGlyGlnAla 20

Qy         160 TTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 219
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 LeuGlyThrThrSerGluPheIleProAsnLeuSerProGluLeuLysLysProLeuSer 40

Qy         220 GAAGGGCAGCCATCATTGAAGAAAATAATACTT 252
               ||||||||||||||||||||||||||||||||
Db          41 GluGlyGlnProSerLeuLysLysIleIleLeu 51

RESULT 9
US-09-174-465D-8
; Sequence 8, Application US/09174465D
; Patent No. 6180364
; GENERAL INFORMATION:
;  APPLICANT:  KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION:  NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
;  FILE REFERENCE:  017753-103
;  CURRENT APPLICATION NUMBER: US/09/174,465D
;  CURRENT FILING DATE:  1998-10-19
;  PRIOR APPLICATION NUMBER:  US 08/482,842
;  PRIOR FILING DATE:  1995-06-07
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 41
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-8

Alignment Scores:
Pred. No.:                4.38e-16        Length:         41
Score:                    193.00          Matches:        33
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              30.39%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-174-465D-8 (1-41)

Qy           1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           9 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 28

Qy          61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTG 99
               |||||||||||||||||||||||||||||||||||||||
Db          29 ProGluLysThrPheThrThrThrProGlyGlyTrpLeu 41

RESULT 10
US-09-599-564A-8
; Sequence 8, Application US/09599564A
; Patent No. 6362318
; GENERAL INFORMATION:
;  APPLICANT: KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
```

```
;  FILE REFERENCE: 017753-127
;  CURRENT APPLICATION NUMBER: US/09/599,564A
;  CURRENT FILING DATE:  2000-06-23
;  PRIOR APPLICATION NUMBER: 09/174,465
;  PRIOR FILING DATE: 1998-10-19
;  PRIOR APPLICATION NUMBER: US 08/482,842
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 41
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-8

Alignment Scores:
Pred. No.:                4.38e-16        Length:         41
Score:                    193.00          Matches:        33
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              30.39%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-599-564A-8 (1-41)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 28

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTG 99
              |||||||||||||||||||||||||||||||||||||||
Db         29 ProGluLysThrPheThrThrThrProGlyGlyTrpLeu 41

RESULT 11
US-09-201-227A-27
; Sequence 27, Application US/09201227A
; Patent No. 6468770
; GENERAL INFORMATION:
;  APPLICANT: Keyes, Linda N.
;  APPLICANT:  Doberstein, Stephen K.
;  APPLICANT:  Buchman, Andrew R.
;  TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF D. MELANOGASTER INSULIN-LIKE
;  TITLE OF INVENTION:  GENES AND USES THEREOF
;  FILE REFERENCE: 7326-066
;  CURRENT APPLICATION NUMBER: US/09/201,227A
;  CURRENT FILING DATE:  1998-11-30
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
;   LENGTH: 41
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-201-227A-27

Alignment Scores:
Pred. No.:                4.38e-16        Length:         41
Score:                    193.00          Matches:        33
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              30.39%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-201-227A-27 (1-41)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          9 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 28

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTG 99
              |||||||||||||||||||||||||||||||||||||||
Db         29 ProGluLysThrPheThrThrThrProGlyGlyTrpLeu 41

RESULT 12
US-09-174-465D-12
; Sequence 12, Application US/09174465D
; Patent No. 6180364
; GENERAL INFORMATION:
;  APPLICANT:  KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION:  NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
;  FILE REFERENCE:  017753-103
;  CURRENT APPLICATION NUMBER: US/09/174,465D
;  CURRENT FILING DATE:  1998-10-19
;  PRIOR APPLICATION NUMBER:  US 08/482,842
;  PRIOR FILING DATE:  1995-06-07
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 30
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-174-465D-12

Alignment Scores:
Pred. No.:                9.24e-13        Length:         30
Score:                    166.00          Matches:        30
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              26.14%          Indels:         0
DB:                       4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-174-465D-12 (1-30)

Qy        253 TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGT 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCysGluValIleCys 20

Qy        313 GACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||
Db         21 AspAspGlyThrSerValLysLeuCysThr 30

RESULT 13
US-09-599-564A-12
; Sequence 12, Application US/09599564A
; Patent No. 6362318
; GENERAL INFORMATION:
;  APPLICANT: KOMAN, Ahment
;  APPLICANT:  CHASSIN, Dorine
;  APPLICANT:  BELLET, Dominique
;  TITLE OF INVENTION: NEW PROTEIN CALLED EPIL/PLACENTIN, PROCESS FOR THE
;  TITLE OF INVENTION:  PREPARATION OF THIS PROTEIN AND PHARMACEUTICAL
;  TITLE OF INVENTION:  COMPOSITION CONTAINING SUCH, DNA CODING FOR SAID
;  TITLE OF INVENTION:  PROTEIN
;  FILE REFERENCE: 017753-127
;  CURRENT APPLICATION NUMBER: US/09/599,564A
;  CURRENT FILING DATE:  2000-06-23
;  PRIOR APPLICATION NUMBER: 09/174,465
;  PRIOR FILING DATE: 1998-10-19
;  PRIOR APPLICATION NUMBER: US 08/482,842
;  PRIOR FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 30
;   TYPE: PRT
;   ORGANISM: Unknown
```

```
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism:EPIL - Early
;   OTHER INFORMATION: Placenta Insulin-Like Peptide
US-09-599-564A-12

Alignment Scores:
Pred. No.:              9.24e-13        Length:         30
Score:                  166.00          Matches:        30
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            26.14%          Indels:         0
DB:                     4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-599-564A-12 (1-30)

Qy        253 TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGT 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCysGluValIleCys 20

Qy        313 GACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||
Db         21 AspAspGlyThrSerValLysLeuCysThr 30

RESULT 14
US-09-201-227A-28
; Sequence 28, Application US/09201227A
; Patent No. 6468770
; GENERAL INFORMATION:
;  APPLICANT: Keyes, Linda N.
;  APPLICANT:  Doberstein, Stephen K.
;  APPLICANT:  Buchman, Andrew R.
;  TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF D. MELANOGASTER INSULIN-LIKE
;  TITLE OF INVENTION:  GENES AND USES THEREOF
;  FILE REFERENCE: 7326-066
;  CURRENT APPLICATION NUMBER: US/09/201,227A
;  CURRENT FILING DATE:  1998-11-30
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
;   LENGTH: 25
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-201-227A-28

Alignment Scores:
Pred. No.:              2.05e-10        Length:         25
Score:                  147.00          Matches:        25
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            23.15%          Indels:         0
DB:                     4               Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x US-09-201-227A-28 (1-25)

Qy        268 AGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCA 327
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 SerGlyArgHisArgPheAspProPheCysCysGluValIleCysAspAspGlyThrSer 20

Qy        328 GTTAAATTATGTACA 342
              |||||||||||||||
Db         21 ValLysLeuCysThr 25

RESULT 15
US-08-443-568B-12
; Sequence 12, Application US/08443568B
; Patent No. 5759807
;  GENERAL INFORMATION:
;    APPLICANT:  Breece, Tim
;    APPLICANT:  Hayenga, Kirk
;    APPLICANT:  Rindersknecht, Ernst
;    APPLICANT:  Vandlen, Richard
;    APPLICANT:  Daniel, Yansura
;    TITLE OF INVENTION:  Process for Producing Relaxin
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/443,568B
;      FILING DATE:  22-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/080,354
;      FILING DATE:  21-JUNE-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Abrams, Samuel B.
;      REGISTRATION NUMBER:  30,605
;      REFERENCE/DOCKET NUMBER:  7842-037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-493-4935
;      TELEFAX:  650-493-5556
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  150 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-443-568B-12

Alignment Scores:
Pred. No.:              6.49e-05        Length:         150
Score:                  105.50          Matches:        43
Percent Similarity:     35.67%          Conservative:   13
Best Local Similarity:  27.39%          Mismatches:     44
Query Match:            16.61%          Indels:         57
DB:                     1               Gaps:           6

US-09-518-842-1_COPY_76_417 (1-342) x US-08-443-568B-12 (1-150)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::     ||||||       :::   |||      ||||||
Db          7 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 26

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                |||:::               ||||||                          ||||||
Db         27 AlaProGln--------------ThrPro----------------------ArgPro 33

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                 |||:::|||   :::      |||||||||   :::   :::         |||
Db         34 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 53

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              ||||||:::   ||||||      |||||||||      |||||||||   ||||||:::
Db         54 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 73

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db         74 LeuProGlnLeuGlnGlnHisValProValLeuLysAspSerSerLeuLeuPheGluGlu 93

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                 ||||||:::|||
Db         94 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 113

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||||||||||||
```

```
Db        114 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 133

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                  ||||||   |||   |||         :::      :::   |||
Db        134 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 150
```

Search completed: June 14, 2003, 19:47:19
Job time : 15.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:46:10 ; Search time 34 Seconds
(without alignments)
2150.892 Million cell updates/sec

Title: US-09-518-842-1_COPY_76_417
Perfect score: 635
Sequence: 1 GCAGAGCTGAGGGGATGTGG.....CTTCAGTTAAATTATGTACA 342

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 817286

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09518842/runat_14062003_175814_12395/app_query.fasta_1.519
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09518842_@CGN_1_1_17_@runat_14062003_175814_12395
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74.5	11.7	1531	12	US-10-007-693-98	Sequence 98, Appl
2	73.5	11.6	269	10	US-09-815-242-5429	Sequence 5429, Ap
3	73.5	11.6	273	10	US-09-815-242-12487	Sequence 12487, A
4	73.5	11.6	273	10	US-09-815-242-12747	Sequence 12747, A
5	73.5	11.6	274	9	US-10-084-205-12	Sequence 12, Appl
6	73.5	11.6	274	10	US-09-925-637-12	Sequence 12, Appl
7	73.5	11.6	890	12	US-10-028-056-3	Sequence 3, Appli
8	71.5	11.3	778	10	US-09-841-132-193	Sequence 193, App
9	71.5	11.3	1530	10	US-09-841-132-178	Sequence 178, App
c 10	70.5	11.8	228	9	US-10-001-254-22	Sequence 22, Appl
c 11	70.5	11.8	514	9	US-10-142-231-68	Sequence 68, Appl
12	69	10.9	741	10	US-09-925-301-930	Sequence 930, App
13	68	10.7	459	9	US-09-759-595-3	Sequence 3, Appli
14	68	10.7	952	9	US-10-216-556-2	Sequence 2, Appli
c 15	67.5	11.2	554	9	US-09-738-626-5066	Sequence 5066, Ap
c 16	66	11.0	133	9	US-09-738-626-5928	Sequence 5928, Ap
c 17	66	11.0	311	9	US-09-804-291-353	Sequence 353, App
c 18	66	11.0	311	10	US-09-886-055-353	Sequence 353, App
c 19	66	11.0	897	9	US-10-099-895-1	Sequence 1, Appli
c 20	66	11.0	2743	9	US-10-037-182-36	Sequence 36, Appl
c 21	66	11.0	3695	9	US-10-037-182-2	Sequence 2, Appli
c 22	65.5	10.9	69	9	US-10-142-231-38	Sequence 38, Appl
23	65.5	10.3	117	10	US-09-815-242-12433	Sequence 12433, A
24	65.5	10.3	117	10	US-09-815-242-12867	Sequence 12867, A
25	65.5	10.3	117	10	US-09-815-242-13094	Sequence 13094, A
26	65	10.2	145	10	US-09-867-550-604	Sequence 604, App
c 27	64	10.7	188	9	US-10-012-542-217	Sequence 217, App
c 28	64	10.7	626	9	US-10-097-340-95	Sequence 95, Appl
29	63.5	10.0	619	12	US-10-000-864-4	Sequence 4, Appli
30	63.5	10.0	2026	10	US-09-801-368-86	Sequence 86, Appl
31	63	9.9	71	10	US-09-948-018-10	Sequence 10, Appl
c 32	63	10.5	173	9	US-09-764-891-4020	Sequence 4020, Ap
c 33	63	10.5	201	9	US-09-924-340-56	Sequence 56, Appl
c 34	63	10.5	201	9	US-09-992-600A-56	Sequence 56, Appl
c 35	63	10.5	201	9	US-10-000-489-56	Sequence 56, Appl
c 36	63	10.5	201	9	US-10-000-986-56	Sequence 56, Appl
37	63	9.9	535	10	US-09-988-200-4	Sequence 4, Appli
38	63	9.9	961	10	US-09-220-091-3	Sequence 3, Appli
c 39	62.5	10.4	213	9	US-09-989-920-206	Sequence 206, App
c 40	62.5	10.4	213	9	US-09-989-920-251	Sequence 251, App
41	62.5	9.8	479	9	US-09-712-363-240	Sequence 240, App
c 42	62.5	10.4	1091	10	US-09-900-237-26	Sequence 26, Appl
c 43	62.5	10.4	1124	10	US-09-771-161A-199	Sequence 199, App
44	62.5	9.8	1421	9	US-10-108-605-323	Sequence 323, App
45	62	9.8	489	10	US-09-956-425-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-007-693-98
; Sequence 98, Application US/10007693
; Patent No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 98
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis serovar D
US-10-007-693-98

Alignment Scores:

Pred. No.:	5.92	Length:	1531
Score:	74.50	Matches:	29
Percent Similarity:	44.79%	Conservative:	14
Best Local Similarity:	30.21%	Mismatches:	26
Query Match:	11.73%	Indels:	27
DB:	12	Gaps:	5

US-09-518-842-1_COPY_76_417 (1-342) x US-10-007-693-98 (1-1531)

```
Qy          4 GAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTG---------------------- 42
              |||   :::|||   |||      ||||||:::::::::
Db        472 GluTyrGlnGlyGlyGlyAlaLeuPheGlyGluAsnIleSerLeuSerGluAsnAlaGly 491

Qy         43 ---CTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACC-----------CCA 87
                 |||::::::               |||||||||   :::
Db        492 ValLeuThrPheLysAspAsnIleValLysThrPheAlaSerAsnGlyLysIleLeuGly 511

Qy         88 GGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 147
              ||||||   :::|||   :::|||:::          |||      :::|||||||
Db        512 GlyGlyAlaIleLeuAlaThrGlyLys-----------ValGluIleThrAsnAsnSer 527

Qy        148 GATGGA----------------------------CAAGCCTTAGGTACGACATCAGAA 177
              :::|||                          |||||||||   |||      |||
Db        528 GluGlyIleSerPheThrGlyAsnAlaArgAlaProGlnAlaLeuProThrGlnGluGlu 547

Qy        178 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 225
              |||   |||      |||   :::      :::|||||||||   |||
Db        548 Phe---ProLeuPheSerLysLysGluGlyArgProLeuSerSerGly 562
```

RESULT 2
US-09-815-242-5429
; Sequence 5429, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5429
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5429

Alignment Scores:

Pred. No.:	4.46	Length:	269
Score:	73.50	Matches:	30
Percent Similarity:	44.23%	Conservative:	16
Best Local Similarity:	28.85%	Mismatches:	30
Query Match:	11.57%	Indels:	28
DB:	10	Gaps:	5

US-09-518-842-1_COPY_76_417 (1-342) x US-09-815-242-5429 (1-269)

```
Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              |||   :::   |||         |||   |||:::   ||||||:::|||
```

```
Db        178 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 197

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              |||::::::   :::      ||||||              ||||||      :::|||:::
Db        198 ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspAspLysProGln 216

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
                                        |||      |||||||||   |||   |||
Db        217 --------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 229

Qy        181 ATTCCTAAT-----------------TTGTCACCAGAG-----------CTGAAGAAA 210
              |||   |||                 ::::::||||||           :::::: ||
Db        230 IleAspAsnTyrLeuGluGlyLysProValThrProGluGluGlnLysValIleGlu-As 249

Qy        211 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 270
              ||||   ::::::   :::|||                                |||||||||::
Db        249 nHisTyrIleArgAsnAlaHis-------------------------LysArgGluLe 260

Qy        271 GGACGTCACA 280
              :      |||
Db        260 uAlaTyrThr 263
```

RESULT 3
US-09-815-242-12487
; Sequence 12487, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12487
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12487

Alignment Scores:

Pred. No.:	4.48	Length:	273
Score:	73.50	Matches:	30
Percent Similarity:	44.23%	Conservative:	16
Best Local Similarity:	28.85%	Mismatches:	30
Query Match:	11.57%	Indels:	28
DB:	10	Gaps:	5

US-09-518-842-1_COPY_76_417 (1-342) x US-09-815-242-12487 (1-273)

```
Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
```

```
                ||| ::: ||| ||| |||::: ||||||:::|||
Db        181 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 200

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
               |||::::::  :::     ||||||          ||||||     :::|||:::
Db        201 ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspAspLysProGln 219

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
                                    |||     |||||||||  |||  |||
Db        220 --------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 232

Qy        181 ATTCCTAAT-----------------TTGTCACCAGAG-----------CTGAAGAAA 210
               |||  |||                :::::||||||          :::::: ||
Db        233 IleAspAsnTyrLeuGluGlyLysProValThrProGluGluGlnLysValIleGlu-As 252

Qy        211 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 270
               ||||  ::::::  :::|||                                 |||||||||::
Db        252 nHisTyrIleArgAsnAlaHis-------------------------LysArgGluLe 263

Qy        271 GGACGTCACA 280
               :  |||
Db        263 uAlaTyrThr 266
```

RESULT 4
US-09-815-242-12747
; Sequence 12747, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12747
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12747

Alignment Scores:

Pred. No.:	4.48	Length:	273
Score:	73.50	Matches:	30
Percent Similarity:	44.23%	Conservative:	16
Best Local Similarity:	28.85%	Mismatches:	30
Query Match:	11.57%	Indels:	28
DB:	10	Gaps:	5

US-09-518-842-1_COPY_76_417 (1-342) x US-09-815-242-12747 (1-273)

```
Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
               |||  :::  |||  |||  |||:::  ||||||:::|||
Db        181 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 200

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
               |||::::::  :::     ||||||          ||||||     :::|||:::
Db        201 ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspAspLysProGln 219

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
                                    |||     |||||||||  |||  |||
Db        220 --------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 232

Qy        181 ATTCCTAAT-----------------TTGTCACCAGAG-----------CTGAAGAAA 210
               |||  |||                :::::||||||          :::::: ||
Db        233 IleAspAsnTyrLeuGluGlyLysProValThrProGluGluGlnLysValIleGlu-As 252

Qy        211 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 270
               ||||  ::::::  :::|||                                 |||||||||::
Db        252 nHisTyrIleArgAsnAlaHis-------------------------LysArgGluLe 263

Qy        271 GGACGTCACA 280
               :  |||
Db        263 uAlaTyrThr 266
```

RESULT 5
US-10-084-205-12
; Sequence 12, Application US/10084205
; Publication No. US20030049648A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil
; TITLE OF INVENTION: 37 Staphylococcus aureus Genes and Polypeptides
; FILE REFERENCE: PB515P1
; CURRENT APPLICATION NUMBER: US/10/084,205
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: PCT/US00/23773
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/151,933
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 12
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-084-205-12

Alignment Scores:

Pred. No.:	4.48	Length:	274
Score:	73.50	Matches:	30
Percent Similarity:	44.23%	Conservative:	16
Best Local Similarity:	28.85%	Mismatches:	30
Query Match:	11.57%	Indels:	28
DB:	9	Gaps:	5

US-09-518-842-1_COPY_76_417 (1-342) x US-10-084-205-12 (1-274)

```
Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
               |||  :::  |||  |||  |||:::  ||||||:::|||
Db        182 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 201

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
               |||::::::  :::     ||||||          ||||||     :::|||:::
Db        202 ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspAspLysProGln 220

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
                                    |||     |||||||||  |||  |||
Db        221 --------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 233

Qy        181 ATTCCTAAT-----------------TTGTCACCAGAG-----------CTGAAGAAA 210
               |||  |||                :::::||||||          :::::: ||
Db        234 IleAspAsnTyrLeuGluGlyLysProValThrProGluGluGlnLysValIleGlu-As 253

Qy        211 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 270
```

```
                ||||   ::::::   :::|||                              |||||||||::
Db        253 nHisTyrIleArgAsnAlaHis-------------------------LysArgGluLe 264

Qy        271 GGACGTCACA 280
                :      |||
Db        264 uAlaTyrThr 267
```

RESULT 6
US-09-925-637-12
; Sequence 12, Application US/09925637
; Patent No. US20020103338A1
; GENERAL INFORMATION:
; APPLICANT: Choi
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Polypeptides
; FILE REFERENCE: PB560
; CURRENT APPLICATION NUMBER: US/09/925,637
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/23773
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 60/151,933
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 08/781,986
; PRIOR FILING DATE: 1997-01-03
; PRIOR APPLICATION NUMBER: US 08/956,171
; PRIOR FILING DATE: 1997-10-20
; PRIOR APPLICATION NUMBER: US 60/009,861
; PRIOR FILING DATE: 1996-01-06
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-637-12

Alignment Scores:

Pred. No.:	4.48	Length:	274
Score:	73.50	Matches:	30
Percent Similarity:	44.23%	Conservative:	16
Best Local Similarity:	28.85%	Mismatches:	30
Query Match:	11.57%	Indels:	28
DB:	10	Gaps:	5

US-09-518-842-1_COPY_76_417 (1-342) x US-09-925-637-12 (1-274)

```
Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              |||   :::   |||         |||   |||:::   |||||||:::|||
Db        182 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 201

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              |||::::::   :::      ||||||           ||||||      :::|||:::
Db        202 ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspAspLysProGln 220

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
                                       |||      |||||||||   |||   |||
Db        221 --------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 233

Qy        181 ATTCCTAAT-----------------TTGTCACCAGAG-----------CTGAAGAAA 210
              |||   |||                 ::::::||||||           :::::: ||
Db        234 IleAspAsnTyrLeuGluGlyLysProValThrProGluGluGlnLysValIleGlu-As 253

Qy        211 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 270
              ||||   ::::::   :::|||                            ||||||||::
Db        253 nHisTyrIleArgAsnAlaHis-------------------------LysArgGluLe 264

Qy        271 GGACGTCACA 280
                :      |||
Db        264 uAlaTyrThr 267
```

RESULT 7
US-10-028-056-3
; Sequence 3, Application US/10028056
; Patent No. US20020152483A1
; GENERAL INFORMATION:
; APPLICANT: REUE, KAREN
; APPLICANT: PTERFY, MIKLOS
; TITLE OF INVENTION: A NOVEL GENE ASSOCIATED WITH REGULATION OF ADIPOSITY AND INSUL
; FILE REFERENCE: 407T-898010US
; CURRENT APPLICATION NUMBER: US/10/028,056
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,772
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-056-3

Alignment Scores:

Pred. No.:	6.56	Length:	890
Score:	73.50	Matches:	20
Percent Similarity:	47.06%	Conservative:	12
Best Local Similarity:	29.41%	Mismatches:	35
Query Match:	11.57%	Indels:	1
DB:	12	Gaps:	1

US-09-518-842-1_COPY_76_417 (1-342) x US-10-028-056-3 (1-890)

```
Qy         79 ACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCC 138
              |||      ||||||   ||||||:::      :::|||:::      :::
Db        324 ThrLeuValGlyGlyAlaLeuLeuAspGlnAsnLysProGlnThrGluMetGlnPheVal 343

Qy        139 AACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCA 198
              |||   :::|||   :::   ||||||      :::      :::|||   :::
Db        344 AsnGluGluAspLeuGluThrLeuGlyAlaAlaAlaProLeuLeuProMetIleGlu--- 362

Qy        199 GAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGC 258
              |||||||||   |||   :::           :::      |||
Db        363 GluLeuLysProProSerAlaSerValValGlnThrAlaAsnLysThrAspSerProSer 382

Qy        259 AAAAAGAGAAGTGGACGTCACAGA 282
              :::||||||      |||   |||
Db        383 ArgLysArgAspLysArgSerArg 390
```

RESULT 8
US-09-841-132-193
; Sequence 193, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 193
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-193

Alignment Scores:

Pred. No.:	11	Length:	778
Score:	71.50	Matches:	29
Percent Similarity:	43.75%	Conservative:	13
Best Local Similarity:	30.21%	Mismatches:	27
Query Match:	11.26%	Indels:	27
DB:	10	Gaps:	5

US-09-518-842-1_COPY_76_417 (1-342) x US-09-841-132-193 (1-778)

```
Qy          4 GAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTG---------- 42
              |||   :::|||   |||      ||||||:::::::::
Db        452 GluTyrGlnGlyGlyGlyAlaLeuPheGlyGluAsnIleSerLeuSerGluAsnAlaGly 471

Qy         43 ---CTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACC-----------CCA 87
                |||::::::               ||||||||   :::
Db        472 ValLeuThrPheLysAspAsnIleValLysThrPheAlaSerAsnGlyLysIleLeuGly 491

Qy         88 GGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 147
              ||||||   :::|||   :::|||:::            |||      :::||||||
Db        492 GlyGlyAlaIleLeuAlaThrGlyLys-----------ValGluIleThrAsnAsnSer 507

Qy        148 GATGGA----------------------------CAAGCCTTAGGTACGACATCAGAA 177
                |||                           |||||||||   |||      |||
Db        508 GlyGlyIleSerPheThrGlyAsnAlaArgAlaProGlnAlaLeuProThrGlnGluGlu 527

Qy        178 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 225
              |||   |||      |||   :::      :::||||||||   |||
Db        528 Phe---ProLeuPheSerLysLysGluGlyArgProLeuSerSerGly 542

RESULT 9
US-09-841-132-178
; Sequence 178, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
;  APPLICANT: Bhatia, Ajay
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Probst, Peter
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
;  TITLE OF INVENTION:  DIAGNOSIS OF CHLAMYDIAL INFECTION
;  FILE REFERENCE: 210121.469C8
;  CURRENT APPLICATION NUMBER: US/09/841,132
;  CURRENT FILING DATE:  2001-04-23
;  NUMBER OF SEQ ID NOS: 599
;  SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 178
;   LENGTH: 1530
;   TYPE: PRT
;   ORGANISM: Chlamydia
US-09-841-132-178

Alignment Scores:
Pred. No.:                13.6          Length:          1530
Score:                    71.50         Matches:         29
Percent Similarity:       43.75%        Conservative:    13
Best Local Similarity:    30.21%        Mismatches:      27
Query Match:              11.26%        Indels:          27
DB:                       10            Gaps:            5

US-09-518-842-1_COPY_76_417 (1-342) x US-09-841-132-178 (1-1530)

Qy          4 GAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTG---------- 42
              |||   :::|||   |||      ||||||:::::::::
Db        472 GluTyrGlnGlyGlyGlyAlaLeuPheGlyGluAsnIleSerLeuSerGluAsnAlaGly 491

Qy         43 ---CTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACC-----------CCA 87
                |||::::::               ||||||||   :::
Db        492 ValLeuThrPheLysAspAsnIleValLysThrPheAlaSerAsnGlyLysIleLeuGly 511

Qy         88 GGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 147
              ||||||   :::|||   :::|||:::            |||      :::||||||
Db        512 GlyGlyAlaIleLeuAlaThrGlyLys-----------ValGluIleThrAsnAsnSer 527

Qy        148 GATGGA----------------------------CAAGCCTTAGGTACGACATCAGAA 177
                |||                           |||||||||   |||      |||
Db        528 GlyGlyIleSerPheThrGlyAsnAlaArgAlaProGlnAlaLeuProThrGlnGluGlu 547

Qy        178 TTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGG 225
              |||   |||      |||   :::      :::||||||||   |||
Db        548 Phe---ProLeuPheSerLysLysGluGlyArgProLeuSerSerGly 562

RESULT 10
US-10-001-254-22
; Sequence 22, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, John C.
;  APPLICANT:  Godzik, Adam
;  APPLICANT:  Pawlowski, Krzysztof
;  APPLICANT:  Fiorentino, Loredana
;  APPLICANT:  Lee, Sug Hyung
;  APPLICANT:  Roth, Wilfred
;  APPLICANT:  Stenner-Liewen, Frank
;  TITLE OF INVENTION: No. US20030049702A1el Death Domain Proteins
;  FILE REFERENCE: P-LJ 5037
;  CURRENT APPLICATION NUMBER: US/10/001,254
;  CURRENT FILING DATE:  2001-11-15
;  PRIOR APPLICATION NUMBER: 60/301,889
;  PRIOR FILING DATE: 2001-06-29
;  PRIOR APPLICATION NUMBER: 09/715,893
;  PRIOR FILING DATE: 2000-11-17
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
;   LENGTH: 228
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-001-254-22

Alignment Scores:
Pred. No.:                9.75          Length:          228
Score:                    70.50         Matches:         23
Percent Similarity:       34.74%        Conservative:    10
Best Local Similarity:    24.21%        Mismatches:      35
Query Match:              11.75%        Indels:          27
DB:                       9             Gaps:            3

US-09-518-842-1_COPY_76_417 (1-342) x US-10-001-254-22 (1-228)

Qy        224 CCTTCAGACAGTGGTTTCTTCAGCTCTGGTGACAAATTAGGAATGAATTCTGATGTCGTA 165
              |||:::   |||      |||         :::      |||   :::|||::::::::
Db         36 ProAsnThrSerSerLeuPheProPro---GluProProGlyAlaSerSerAsnIleIle 54

Qy        164 CCTAAGGCTTGTCCATCTTTG---------------------------------------- 144
              |||      |||      |||
Db         55 ProValTyrCysAlaLeuLeuAlaThrValIleLeuGlyLeuLeuAlaTyrValAlaPhe 74

Qy        143 --------------------------------TTGTTGGAGGTTGACACCATTTCTTTG 117
                                             |||   :::      |||:::   |||
Db         75 LysCysTrpArgSerHisLysGlnArgGlnGlnLeuAlaLysAlaArgThrValGluLeu 94

Qy        116 GGACGTCCAGATTCCAGCAGCCACCCTCCTGGGGTGGTGGTGAATGTCTTCTCAGGCATG 57
              |||   ||||||                                 |||||||||
Db         95 GlyAspProAspArgAspGlnArgArgGlyAsp------SerAsnValPheValAspSer 112

Qy         56 GGGCAATATGACAGCAAGTGTTTTCCAAATCGGGGACCACATCCC 12
                              |||   |||::::::||||||||||||
Db        113 ProProSerLeuGluProCysIleProSerGlnGlyProHisPro 127

RESULT 11
US-10-142-231-68
; Sequence 68, Application US/10142231
; Publication No. US20030077796A1
; GENERAL INFORMATION:
;  APPLICANT: Croteau, Rodney et al.
;  TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
;  FILE REFERENCE: 62773
;  CURRENT APPLICATION NUMBER: US/10/142,231
;  CURRENT FILING DATE:  2002-05-08
;  PRIOR APPLICATION NUMBER: 60/165,250
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 95
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
;   LENGTH: 514
```

```
;   TYPE: PRT
;   ORGANISM: Taxus cuspidata
US-10-142-231-68

Alignment Scores:
Pred. No.:              12.7          Length:          514
Score:                  70.50         Matches:         24
Percent Similarity:     47.95%        Conservative:    11
Best Local Similarity:  32.88%        Mismatches:      25
Query Match:            11.75%        Indels:          13
DB:                     9             Gaps:            4

US-09-518-842-1_COPY_76_417 (1-342) x US-10-142-231-68 (1-514)

Qy        209 TTCTTCAGCTCTGGT---------GACAAATTAGGAATGAATTCTGATGTCGTACCT--- 162
              |||      ||||||         |||      |||         :::::::::|||
Db        428 PheLeuLysSerGlyLysGluValAspValLysGlyArgGluPheGluLeuIleProPhe 447

Qy        161 ------------AAGGCTTGTCCA------TCTTTGTTGTTGGAGGTTGACACCATTTCT 120
                      :::  ||||||      ||||||   |||   |||   |||   :::
Db        448 GlySerGlyArgArgMetCysProGlyMetSerLeuAlaLeuSerValValThrTyrThr 467

Qy        119 TTGGGACGTCCAGATTCCAGC---------AGCCACCCTCCTGGGGTGGTGGTGAATGTC 69
              |||||||||         |||         |||   |||   |||:::::::::::::
Db        468 LeuGlyArgLeuLeuGlnSerPheGluTrpSerValProGluGlyMetIleIleAspMet 487

Qy         68 TTCTCAGGCATGGGGCAATATGACAGCAAGTGTTTTCCA 30
                    |||:::|||             |||      |||
Db        488 ThrGluGlyLeuGlyLeuThrMetProLysAlaValPro 500

RESULT 12
US-09-925-301-930
; Sequence 930, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 930
;   LENGTH: 741
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (8)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (19)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (282)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-930

Alignment Scores:
Pred. No.:              21.7          Length:          741
Score:                  69.00         Matches:         18
Percent Similarity:     50.00%        Conservative:    6
Best Local Similarity:  37.50%        Mismatches:      22
Query Match:            10.87%        Indels:          2
DB:                     10            Gaps:            1

US-09-518-842-1_COPY_76_417 (1-342) x US-09-925-301-930 (1-741)

Qy         55 CCCATGCCTGAGAAGACATTCACCACCACCCCA------GGAGGGTGGCTGCTGGAATCT 108
                 |||:::|||   |||:::   |||      |||         |||         |||:::
Db         39 ProIleProArgLysSerAlaThrValIleProGluIleGluGlyIleLysAlaGluAla 58

Qy        109 GGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACG 168
                          :::|||      :::|||         ||||||||||||:::
Db         59 LysAlaLeuAspAspMetPheGluSerSerThrLeuSerAspGlyGlnAlaIleAlaAsp 78

Qy        169 ACATCAGAATTCATTCCTAATTTG 192
                 ||||||   ||||||   |||
Db         79 GlnSerGluIleIleProThrLeu 86

RESULT 13
US-09-759-595-3
; Sequence 3, Application US/09759595
; Publication No. US20030059916A1
; GENERAL INFORMATION:
;  APPLICANT: Wesche, Holger
;  APPLICANT:  Li, Shyun
;  APPLICANT:  Tularik Inc.
;  TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
;  FILE REFERENCE: 018781-003910US
;  CURRENT APPLICATION NUMBER: US/09/759,595
;  CURRENT FILING DATE:  2001-01-13
;  PRIOR APPLICATION NUMBER: US 60/176,395
;  PRIOR FILING DATE: 2000-01-13
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 459
;   TYPE: PRT
;   ORGANISM: Mus sp.
US-09-759-595-3

Alignment Scores:
Pred. No.:              24.5          Length:          459
Score:                  68.00         Matches:         21
Percent Similarity:     47.22%        Conservative:    13
Best Local Similarity:  29.17%        Mismatches:      20
Query Match:            10.71%        Indels:          18
DB:                     9             Gaps:            2

US-09-518-842-1_COPY_76_417 (1-342) x US-09-759-595-3 (1-459)

Qy        166 ACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT------ 219
              |||   |||   :::|||   ||||||:::      :::   :::   ||||||
Db          6 ThrProSerThrTyrIleArgAsnLeuAsnValGlyIleLeuArgLysLeuSerAspPhe 25

Qy        220 ---GAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGT 276
                 :::   |||         ||||||:::   :::::   ||||||   ||||||
Db         26 IleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAsp 45

Qy        277 CACAGATTTGATCCATTC---------------------------------------- 294
                 |||::::::   |||
Db         46 AspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSer 65

Qy        295 ---TGTTGTGAAGTAATTTGTGACGATGGAACTTCA 327
                    ||||||::::::   |||   ||||||:::
Db         66 ProThrCysGluLeuLeuPheAspTrpGlyThrThr 77

RESULT 14
US-10-216-556-2
; Sequence 2, Application US/10216556
; Publication No. US20030073113A1
; GENERAL INFORMATION:
;  APPLICANT: Van Houten, Bennet
;  APPLICANT:  Skorvaga, Milan
;  TITLE OF INVENTION: THERMOSTABLE UvrA AND UvrB POLYPEPTIDES AND METHODS OF USE
;  FILE REFERENCE: 265.00330101
;  CURRENT APPLICATION NUMBER: US/10/216,556
;  CURRENT FILING DATE:  2002-08-10
;  PRIOR APPLICATION NUMBER: 60/311,336
;  PRIOR FILING DATE: 2001-08-10
;  NUMBER OF SEQ ID NOS: 44
```

```
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 952
;   TYPE: PRT
;   ORGANISM: BACILLUS CALDOTENAX
US-10-216-556-2
```

Alignment Scores:

Pred. No.:	31	Length:	952
Score:	68.00	Matches:	20
Percent Similarity:	46.15%	Conservative:	10
Best Local Similarity:	30.77%	Mismatches:	17
Query Match:	10.71%	Indels:	18
DB:	9	Gaps:	3

US-09-518-842-1_COPY_76_417 (1-342) x US-10-216-556-2 (1-952)

```
Qy         19 GGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACC 78
              ||||||      |||   |||
Db        559 GlyProGlyAlaGlyIleHis---------------------------------------- 565

Qy         79 ACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCC 138
                 ||||||   ::::::   :::|||   |||:::|||::::::
Db        566 ---------GlyGlyGluValIleSerAlaGlyThrProGluGluValMetGluAspPro 582

Qy        139 AACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCA 198
              |||:::      |||      |||   :::      :::||||||||||   |||   |||
Db        583 AsnSerLeuThrGlySerTyrLeu---SerGlyLysLysPheIlePro---LeuProPro 600

Qy        199 GAGCTGAAGAAACCA 213
              |||   :::||||||
Db        601 GluArgArgLysPro 605
```

```
RESULT 15
US-09-738-626-5066
; Sequence 5066, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5066
;   LENGTH: 554
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-5066
```

Alignment Scores:

Pred. No.:	30	Length:	554
Score:	67.50	Matches:	16
Percent Similarity:	51.79%	Conservative:	13
Best Local Similarity:	28.57%	Mismatches:	24
Query Match:	11.25%	Indels:	3
DB:	9	Gaps:	2

US-09-518-842-1_COPY_76_417 (1-342) x US-09-738-626-5066 (1-554)

```
Qy        197 GGTGACAAATTAGGAATGAATTCTGATGTCGTACCTAAGGCTTGTCCATCTTTGTTGTTG 138
              |||:::   ||||||          |||:::|||                 :::
Db        113 GlyGluTyrLeuGlyLysThrValGlnValIleProHisIleThrAspGluIleLysAla 132

Qy        137 GAGGTTGACACCATTTCTTTGGGACGTCCAGATTCCAGCAGCCACCCTCCTGGGGTGGTG 78
                 :::      :::|||:::|||   ||||||:::      :::   |||   ||||||
Db        133 ArgIle------LeuSerMetGlyGluProAspAlaHisGlyAsnAlaProAspValVal 150

Qy         77 GTGAATGTCTTCTCAGGC---ATGGGGCAATATGACAGCAAGTGTTTT 33
              ::::::         |||   :::|||      :::|||:::   |||
Db        151 IleSerGluValGlyGlyThrValGlyAspIleGluSerGlnProPhe 166
```

```
Search completed: June 14, 2003, 19:58:26
Job time : 39 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:30:20 ; Search time 34 Seconds
(without alignments)
2680.689 Million cell updates/sec

Title: US-09-518-842-1_COPY_76_417
Perfect score: 635
Sequence: 1 GCAGAGCTGAGGGGATGTGG.....CTTCAGTTAAATTATGTACA 342

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09518842/runat_14062003_175810_12235/app_query.fasta_1.519
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09518842_@CGN_1_1_114_@runat_14062003_175810_12235 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		619	97.5	124	19	AAW69169	Zinsl NF protein.
2		619	97.5	139	17	AAR89134	Human early placen
3		619	97.5	139	18	AAW17676	Human relaxin-rela
4		619	97.5	139	19	AAW69168	Zinsl protein. Ho
5		619	97.5	139	20	AAY26926	Human insulin-like
6		619	97.5	139	20	AAW99574	Human early placen
7		619	97.5	159	19	AAW69170	N-terminally tagge
8		142	22.4	182	5	AAP40156	Sequence of porcin
9		141	22.2	182	4	AAP30392	Sequence of porcin
10		111	17.5	178	14	AAR31958	Squamous cell spec
11		105.5	16.6	150	16	AAR64904	Prolrelaxin fragme
12		105.5	16.6	162	11	AAR07987	H2 prorelaxin gene
13		105.5	16.6	185	10	AAP94621	Amino acid sequenc
14		103.5	16.3	185	5	AAP40154	Sequence of human
15		103.5	16.3	185	5	AAP40155	Sequence of human
16		103.5	16.3	185	10	AAP94622	Amino acid sequenc
17		103.5	16.3	185	23	ABG61819	Prostate cancer-as
18		103	16.2	185	5	AAP40108	Sequence of human
c 19		103	17.2	787	22	ABG05453	Novel human diagno
c 20		103	17.2	787	22	ABG08971	Novel human diagno
21		99.5	15.7	164	11	AAR07988	H2 prorelaxin dedu
22		97	15.3	220	22	AAU02911	Angiotensin conver
23		78	12.3	927	22	ABG25847	Novel human diagno
24		75.5	11.9	612	22	ABG11791	Novel human diagno
25		74.5	11.7	1531	22	AAU38904	C. trachomatis CT8
26		73.5	11.6	269	22	AAU33933	Staphylococcus aur
27		73.5	11.6	273	22	AAU36894	Staphylococcus aur
28		73.5	11.6	273	22	AAU37154	Staphylococcus aur
29		73.5	11.6	274	22	AAU00833	S. aureus NH3-depe
30		73.5	11.6	282	22	AAG67271	Amino acid sequenc
c 31		73.5	12.2	402	21	AAG29442	Arabidopsis thalia
c 32		73	12.2	210	22	AAU49608	Propionibacterium
33		73	11.5	545	23	ABB89989	Human polypeptide
34		73	11.5	636	22	AAB94352	Human protein sequ
35		73	11.5	747	22	AAM40070	Human polypeptide
36		73	11.5	747	22	AAB93124	Human protein sequ
c 37		73	12.2	994	18	AAW27147	HMG-CoA reductase
c 38		72.5	12.1	1754	22	ABG18818	Novel human diagno
c 39		72.5	12.1	1759	22	AAO09273	Human polypeptide
40		72	11.3	732	20	AAY42697	Mouse serine-threo
41		72	11.3	806	22	AAB65622	Novel protein kina
42		71.5	11.3	647	23	ABP26017	Streptococcus poly
43		71.5	11.3	778	21	AAB13642	C. trachomatis pmp
44		71.5	11.3	778	22	AAG83210	Protein encoded by
45		71.5	11.3	778	23	ABB94181	Chlamydia protein

ALIGNMENTS

RESULT 1
AAW69169
ID AAW69169 standard; Protein; 124 AA.
XX
AC AAW69169;
XX
DT 07-OCT-1998 (first entry)
XX
DE Zinsl NF protein.
XX
KW Zinsl; human; placenta; placentin; pancreatic islet cell proliferation;
KW insulin secretion; diabetes; therapy; Zinsl NF.
XX
OS Homo sapiens.
XX
PN WO9827210-A1.
XX
PD 25-JUN-1998.
XX

```
PF   16-DEC-1997;   97WO-US23326.
XX
PR   16-DEC-1996;   96US-0033003.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Hoffmann RC,  Humes JM,  Jaspers SR,  Ren HP;
PI   Sprugel KH;
XX
DR   WPI; 1998-362779/31.
XX
PT   New isolated protein, Zins1 - obtained from human placenta, which
PT   increases the proliferation of pancreatic islet cells, used for
PT   treating diabetes
XX
PS   Example 3; Page 64; 77pp; English.
XX
CC   This sequence is the human Zins1 NF protein of the invention. The Zins1
CC   protein was isolated form human placenta, and is believed to be a new
CC   version of the mature protein of placentin, having disulphide bonded A
CC   and B chains. The protein can be used for stimulating the proliferation
CC   of pancreatic islets to increase insulin secretory capacity of mammals.
CC   In particular it can be used for the treatment of diabetes. It can also
CC   be used for stimulating in vitro proliferation of pancreatic islet cells.
CC   It can also be used for production of antibodies and in detection and
CC   diagnosis.
XX
SQ   Sequence   124 AA;

Alignment Scores:
Pred. No.:                3.7e-68       Length:          124
Score:                    619.00        Matches:         114
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              97.48%        Indels:          0
DB:                       19            Gaps:            0

US-09-518-842-1_COPY_76_417 (1-342) x AAW69169 (1-124)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 30

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 50

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 70

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 90

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 110

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              ||||||||||||||||||||||||||||||||||||||||||
Db        111 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 124

RESULT 2
AAR89134
ID   AAR89134 standard; Protein; 139 AA.
XX
AC   AAR89134;
XX
DT   31-JUL-1996  (first entry)
XX
DE   Human early placental insulin-like protein.
XX
KW   Insulin growth hormone family; early placental insulin like protein;
KW   placentin; placental tissue; cytotrophoblast; trimester; probe; primer;
KW   amplification;  polymerase chain reaction; tyrosine phosphorylation;
KW   cellular protein; growth factor; human; lactation; promoter; PCR;
KW   regeneration; nerve; muscle; skin; bone tissue.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..17
FT                   /note= "signal peptide"
FT   Peptide         18..58
FT                   /note= "B-chain peptide"
FT   Peptide         59..109
FT                   /note= "C-peptide, links B and A chains in pre-E"
FT   Peptide         110..139
FT                   /note= "A-chain peptide"
XX
PN   WO9534653-A1.
XX
PD   21-DEC-1995.
XX
PF   12-JUN-1995;   95WO-FR00766.
XX
PR   13-JUN-1994;   94FR-0007191.
XX
PA   (INSR ) INST ROUSSY GUSTAVE.
XX
PI   Bellet D,  Chassin D,  Koman A;
XX
DR   WPI; 1996-049682/05.
DR   N-PSDB; AAT10275.
XX
PT   Early placental insulin-like protein, EPIL/placentin - contains
PT   growth factor-like activity useful for e.g. promoting lactation or
PT   for regeneration of nerve, muscle, skin or bone tissue
XX
PS   Claim 1; Page 14-15; 25pp; French.
XX
CC   This is the amino acid sequence of a novel member of the insulin growth
CC   hormone family designated early placental insulin like (EPIL) protein or
CC   placentin.  The encoding gene has been found to be expressed exclusively
CC   in the early placental tissue with a small amount in full term placental
CC   tissue.  No other tissue expresses this gene. The gene was obtained from
CC   a cDNA library derived from mRNA isolated from cytotrophoblasts derived
CC   from first trimester placental tissue using, as a probe, a fragment of
CC   the gene amplified by primers AAT10276-7.  Although the specific
CC   activity of the protein remains to be elucidated, it is thought that the
CC   protein will induce tyrosine phosphorylation of cellular proteins and may
CC   have growth factor-like activities e.g. human growth factor type 1 or
CC   lactation promoter activities. It may also be used to regenerate e.g.
CC   nerve, muscle, skin or bone tissue.
XX
SQ   Sequence   139 AA;

Alignment Scores:
Pred. No.:                3.85e-68      Length:          139
Score:                    619.00        Matches:         114
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              97.48%        Indels:          0
DB:                       17            Gaps:            0

US-09-518-842-1_COPY_76_417 (1-342) x AAR89134 (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65
```

```
Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              ||||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139
```

RESULT 3
AAW17676
ID AAW17676 standard; Protein; 139 AA.
XX
AC AAW17676;
XX
DT 24-JUL-1997 (first entry)
XX
DE Human relaxin-related factor-2 (RRF-2).
XX
KW Relaxin-related factor-2; RRF-2; testis; sperm; growth factor;
KW infertility.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..26
FT /label= Sig_peptide
FT Misc-difference 124
FT /note= "conserved Cys residue indicative of
FT insulin family member"
FT Misc-difference 125
FT /note= "conserved Cys residue indicative of
FT insulin family member"
FT Misc-difference 129
FT /note= "conserved Cys residue indicative of
FT insulin family member"
FT Misc-difference 138
FT /note= "conserved Cys residue indicative of
FT insulin family member"
XX
PN WO9716549-A2.
XX
PD 09-MAY-1997.
XX
PF 01-NOV-1996; 96WO-US17342.
XX
PR 21-FEB-1996; 96US-0012016.
PR 03-NOV-1995; 95US-0006221.
XX
PA (REGE-) REGENERON PHARM INC.
XX
PI Davis S;
XX
DR WPI; 1997-272118/24.
DR N-PSDB; AAT68419.
XX
PT New isolated relaxin-related factor genes - used to develop products
PT which can be used in diagnosis and therapy, e.g. in fertility and
PT pregnancy applications
XX
PS Example 3; Fig 4; 34pp; English.
XX
CC Human relaxin-related factor-2 (RRF-2) (AAW17676) is a placenta-
CC specific growth factor related to relaxin and to the insulin family
CC of ligands. RRF-2 cDNA (AAT68419) was isolated in a search of
CC expressed sequence tags for sequences related to relaxin. RRF-2
CC displays all the expected features of a new insulin family member,
CC partic. with regard to a cluster of four cysteine residues at the
CC C-terminus of the molecule. A related testis-specific factor,
CC RRF-1 (AAW17675), has also been identified. RRF-2 can be produced
CC in transformed host cells for use in the prepn. of antibodies and
CC therapeutic compsns., or as a growth factor for maintaining cells
CC in culture. RRF-2 may be useful for modulating the reproductive
CC physiology of mammals during pregnancy and parturition.
XX
SQ Sequence 139 AA;

Alignment Scores:

Pred. No.:	3.85e-68	Length:	139
Score:	619.00	Matches:	114
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.48%	Indels:	0
DB:	18	Gaps:	0

US-09-518-842-1_COPY_76_417 (1-342) x AAW17676 (1-139)

```
Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              ||||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139
```

RESULT 4
AAW69168
ID AAW69168 standard; Protein; 139 AA.
XX
AC AAW69168;
XX
DT 07-OCT-1998 (first entry)
XX
DE Zins1 protein.
XX
KW Zins1; human; placenta; placentin; pancreatic islet cell proliferation;
KW insulin secretion; diabetes; therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 26..43
FT /note= "B chain of Zins1"
FT Region 115..0
FT /note= "A chain of Zins1"
XX
PN WO9827210-A1.
XX
PD 25-JUN-1998.
XX
PF 16-DEC-1997; 97WO-US23326.
XX
PR 16-DEC-1996; 96US-0033003.

```
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Hoffmann RC,  Humes JM,  Jaspers SR,  Ren HP;
PI   Sprugel KH;
XX
DR   WPI; 1998-362779/31.
DR   N-PSDB; AAV44663.
XX
PT   New isolated protein, Zins1 - obtained from human placenta, which
PT   increases the proliferation of pancreatic islet cells, used for
PT   treating diabetes
XX
PS   Claim 4; Page 60-61; 77pp; English.
XX
CC   This sequence is the human Zins1 protein of the invention. The Zins1
CC   protein was isolated form human placenta, and is believed to be a new
CC   version of the mature protein of placentin, having disulphide bonded A
CC   and B chains. The protein can be used for stimulating the proliferation
CC   of pancreatic islets to increase insulin secretory capacity of mammals.
CC   In particular it can be used for the treatment of diabetes. It can also
CC   be used for stimulating in vitro proliferation of pancreatic islet cells.
CC   It can also be used for production of antibodies and in detection and
CC   diagnosis.
XX
SQ   Sequence   139 AA;

Alignment Scores:
Pred. No.:              3.85e-68        Length:         139
Score:                  619.00          Matches:        114
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            97.48%          Indels:         0
DB:                     19              Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x AAW69168 (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 5
AAY26926
ID   AAY26926 standard; Protein; 139 AA.
XX
AC   AAY26926;
XX
DT   21-DEC-1999  (first entry)
XX
DE   Human insulin-like 4 protein.
XX
KW   Human; insulin-like 4; INSL4; embryonic; c bone tissue; ligament;
KW   early-placental insulin-like protein; EPIL; antibody; probe; primer;
KW   diagnosis; pathology; differentiation; proliferation; cartilage;
KW   ossification; osteoporosis; dysplasia.
XX
OS   Homo sapiens.
XX
PN   WO9937780-A1.
XX
PD   29-JUL-1999.
XX
PF   22-JAN-1999;   99WO-FR00137.
XX
PR   23-JAN-1998;   98FR-0000715.
XX
PA   (INSR ) INST ROUSSY GUSTAVE.
XX
PI   Laurent A,  Bellet D;
XX
DR   WPI; 1999-469135/39.
DR   N-PSDB; AAX83561.
XX
PT   Expression of the INSL4 gene in human embryonic bone tissue and
PT   ligaments,  -
XX
PS   Disclosure; Fig 1; 60pp; French.
XX
CC   This sequence represents the human protein encoded by the insulin-like 4
CC   (INSL4) gene which is designated early-placental insulin-like (EPIL)
CC   protein. The INSL4 gene is expressed in human embryonic bone tissue and
CC   ligaments and encodes 3 different EPIL proteins designated EPIL 1, 2 or
CC   3. EPIL 1 is a single chain comprising amino acids 18-139, EPIL 2 is a
CC   2 chain protein with chain A comprising amino acids 115-139 and chain B
CC   comprising amino acids 18-58 and EPIL 3 is a single chain comprising
CC   amino acids 59-114. The nucleic acids and protein or antibodies against
CC   EPIL 1, 2 or 3, and probes or primers for INSL4 are useful for the
CC   diagnosis of pathology associated with abnormal differentiation and/or
CC   proliferation of bone tissue or ligaments or development of abnormal
CC   cartilage and/or abnormal ossification of forming bones, e.g.
CC   osteoporosis or dysplasia. Compositions capable of modulation
CC   differentiation, regeneration and/or proliferation of bone tissue
CC   and/or ligament cells are also useful in treating bone disease.
XX
SQ   Sequence   139 AA;

Alignment Scores:
Pred. No.:              3.85e-68        Length:         139
Score:                  619.00          Matches:        114
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            97.48%          Indels:         0
DB:                     20              Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x AAY26926 (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125
```

```
Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 6
AAW99574
ID   AAW99574 standard; Protein; 139 AA.
XX
AC   AAW99574;
XX
DT   22-JUN-1999  (first entry)
XX
DE   Human early placental insulin-like (EPIL) polypeptide.
XX
KW   INSL-4; insulin-like gene; EPIL; early placental insulin-like; antibody;
KW   vector; probe; hybridisation; tumour; hypoglycaemia; hyperglycaemia;
KW   diabetes; cardiovascular.
XX
OS   Homo sapiens.
XX
PN   WO9909172-A1.
XX
PD   25-FEB-1999.
XX
PF   12-AUG-1998;   98WO-FR01799.
XX
PR   03-NOV-1997;   97FR-0013802.
PR   14-AUG-1997;   97FR-0010387.
XX
PA   (INSR ) INST ROUSSY GUSTAVE.
XX
PI   Bellet D,  Bidart JM,  Troalen F,  Mock P;
XX
DR   WPI; 1999-181038/15.
DR   N-PSDB; AAX27490.
XX
PT   EPIL polypeptides encoded by insulin-like gene 4 - and corresponding
PT   nucleic acids, antibodies, probes, primers, etc.
XX
PS   Claim 1; Fig 1; 119pp; French.
XX
CC   This sequence represent the early placental insulin-like (EPIL) protein
CC   encoded by the INSL-4 (insulin-like gene 4) gene. The polypeptide,
CC   antibodies to the polypeptide, vectors containing the coding sequence and
CC   probes derived from the coding sequence, can be used to treat tumours,
CC   preferably angioproliferative tumours, especially Kaposi's sarcoma,
CC   tumours of the pancreas, liver, uterus or breast, angiosarcomas,
CC   glioblastomas, neuroblastomas, rhabdomyosarcomas or leiomyosarcomas;
CC   to promote vascularisation of specific tissues; to treat retinopathy,
CC   macular degeneration, psoriasis, endometriosis, rheumatoid arthritis,
CC   atherosclerosis or hyperthyroidism; to treat post-angioplastic
CC   restenosis; to promote or inhibit embryo implantation; to prevent and/or
CC   treat disorders directly or indirectly connected with insulin-like
CC   activity; to prevent and/or treat disorders directly or indirectly
CC   connected with a dysfunction in carbohydrate metabolism, especially
CC   connected with hypo glycaemia or hyperglycaemia, especially gestational
CC   diabetes and diabetic complications, especially cardiovascular
CC   complications.
XX
SQ   Sequence   139 AA;

Alignment Scores:
Pred. No.:              3.85e-68        Length:         139
Score:                  619.00          Matches:        114
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            97.48%          Indels:         0
DB:                     20              Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x AAW99574 (1-139)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         26 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 45

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         46 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 65

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         66 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 85

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         86 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 105

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        106 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 125

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||||
Db        126 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 139

RESULT 7
AAW69170
ID   AAW69170 standard; Protein; 159 AA.
XX
AC   AAW69170;
XX
DT   07-OCT-1998  (first entry)
XX
DE   N-terminally tagged Zinsl protein.
XX
KW   Zinsl; human; placenta; placentin; pancreatic islet cell proliferation;
KW   insulin secretion; diabetes; therapy.
XX
OS   Homo sapiens.
XX
PN   WO9827210-A1.
XX
PD   25-JUN-1998.
XX
PF   16-DEC-1997;   97WO-US23326.
XX
PR   16-DEC-1996;   96US-0033003.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Hoffmann RC,  Humes JM,  Jaspers SR,  Ren HP;
PI   Sprugel KH;
XX
DR   WPI; 1998-362779/31.
DR   N-PSDB; AAV44664.
XX
PT   New isolated protein, Zinsl - obtained from human placenta, which
PT   increases the proliferation of pancreatic islet cells, used for
PT   treating diabetes
XX
PS   Example 1; Page 62-63; 77pp; English.
XX
CC   This sequence is a N-terminally tagged version of the human Zinsl
CC   protein of the invention. The Zinsl protein was isolated form human
CC   placenta, and is believed to be a new version of the mature protein of
CC   placentin, having disulphide bonded A and B chains. The protein can be
CC   used for stimulating the proliferation of pancreatic islets to increase
CC   insulin secretory capacity of mammals. In particular it can be used for
CC   the treatment of diabetes. It can also be used for stimulating in vitro
CC   proliferation of pancreatic islet cells. It can also be used for
CC   production of antibodies and in detection and diagnosis.
XX
SQ   Sequence   159 AA;

Alignment Scores:
Pred. No.:              4.05e-68        Length:         159
```

```
Score:                  619.00          Matches:        114
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            97.48%          Indels:         0
DB:                     19              Gaps:           0

US-09-518-842-1_COPY_76_417 (1-342) x AAW69170 (1-159)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 AlaGluLeuArgGlyCysGlyProArgPheGlyLysHisLeuLeuSerTyrCysProMet 65

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ProGluLysThrPheThrThrThrProGlyGlyTrpLeuLeuGluSerGlyArgProLys 85

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 GluMetValSerThrSerAsnAsnLysAspGlyGlnAlaLeuGlyThrThrSerGluPhe 105

Qy        181 ATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 IleProAsnLeuSerProGluLeuLysLysProLeuSerGluGlyGlnProSerLeuLys 125

Qy        241 AAAATAATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 LysIleIleLeuSerArgLysLysArgSerGlyArgHisArgPheAspProPheCysCys 145

Qy        301 GAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGTACA 342
              |||||||||||||||||||||||||||||||||||||||||
Db        146 GluValIleCysAspAspGlyThrSerValLysLeuCysThr 159

RESULT 8
AAP40156
ID   AAP40156 standard; Protein; 182 AA.
XX
AC   AAP40156;
XX
DT   11-FEB-1992  (first entry)
XX
DE   Sequence of porcine preprorelaxin.
XX
KW   Labour; birth; hormone; relaxin.
XX
OS   Pig.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /label= signal
FT   Peptide         26..57
FT                   /label= B-chain
FT   Peptide         58..160
FT                   /label= C-peptide
FT   Peptide         161..182
FT                   /label= A-chain
XX
PN   EP101309-A.
XX
PD   22-FEB-1984.
XX
PF   11-AUG-1983;   83EP-0304662.
XX
PR   12-AUG-1982;   82AU-0005352.
PR   11-AUG-1983;   83AU-0017906.
PR   01-JAN-1988;   88EP-0104503.
XX
PA   (FLOR-) HOWARD FLOREY INST.
PA   (HOWA-) HOWARD FLOREY INST.
XX
PI   Hudson PJ,  Shine J,  Niall HD,  Tregear GW;
XX
DR   WPI; 1984-050918/09.
DR   N-PSDB; AAN40125, AAN40126.
XX
PT   Genes for human relaxin, prorelaxin and preprorelaxin prodn. -
PT   prepd. by recombinant DNA techniques
XX
PS   Disclosure; Fig 3; 51pp; English.
XX
CC   The inventors claim the gene for the expression of human
CC   preprorelaxin (HPP) and its sub-units. Also claimed are synthetic
CC   HPP, HP, fragments and analogues. Human relaxin and its analogues
CC   are prepd. for therapeutic purposes, esp. in clinical intervention
CC   in cases of difficult labour.
XX
SQ   Sequence   182 AA;

Alignment Scores:
Pred. No.:              1.01e-08        Length:         182
Score:                  142.00          Matches:        41
Percent Similarity:     40.00%          Conservative:   7
Best Local Similarity:  34.17%          Mismatches:     32
Query Match:            22.36%          Indels:         40
DB:                     5               Gaps:           2

US-09-518-842-1_COPY_76_417 (1-342) x AAP40156 (1-182)

Qy        100 CTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCC 159
              ||||||:::|||   |||   |||   :::   :::|||      ||||||   :::
Db         63 LeuGluThrGlyProProAlaGluThrMetProSerSerIleThrLysAspAlaGluIle 82

Qy        160 TTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 219
              |||            ||||||:::|||||||||      |||||||||      ||||||
Db         83 LeuLysMetMetLeuGluPheValProAsnLeuProGlnGluLeuLysAlaThrLeuSer 102

Qy        220 GAAGGGCAGCCATCATTG---------------------------------------- 237
              |||   ||||||||||||
Db        103 GluArgGlnProSerLeuArgGluLeuGlnGlnSerAlaSerLysAspSerAsnLeuAsn 122

Qy        238 ------------AAGAAAATAATACTT------------------------------ 252
                         ||||||||||||||||
Db        123 PheGluGluPheLysLysIleIleLeuAsnArgGlnAsnGluAlaGluAspLysSerLeu 142

Qy        253 ------------------------------TCCCGCAAAAAGAGAAGTGGACGTCAC 279
                                           ||||||||||||||||      |||
Db        143 LeuGluLeuLysAsnLeuGlyLeuAspLysHisSerArgLysLysArgLeuPheArgSer 162

Qy        280 AGATTTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                           ||||||:::|||   |||                  :::||||||
Db        163 ThrLeuSerGluLysCysCysGlnValGlyCysIleArgLysAspIleAlaArgLeuCys 182

RESULT 9
AAP30392
ID   AAP30392 standard; Protein; 182 AA.
XX
AC   AAP30392;
XX
DT   25-MAY-1992  (first entry)
XX
DE   Sequence of porcine preprorelaxin.
XX
KW   Relaxin; hormone.
XX
OS   Sus scrofa domestica.
XX
PN   EP86649-A.
XX
PD   24-AUG-1983.
XX
PF   12-FEB-1983;   83EP-0021967.
XX
PR   12-FEB-1982;   82AU-0002695.
PR   11-FEB-1983;   83AU-0011834.
XX
PA   (FLOR-) FLOREY INST EXP PHY.
PA   (HOWA-) HOWARD FLOREY INST.
```

```
XX
PI   Hudson PJ,  Haley JD,  Niall HD,  Shine J;
XX
DR   WPI; 1983-748587/35.
DR   N-PSDB; AAN30196.
XX
PT   Genes and DNA transfer vectors for prorelaxin expression - useful
PT   in prodn. of porcine relaxin for veterinary and human use
XX
PS   Disclosure; Fig 5; 50pp; English.
XX
CC   The inventors claim synthetic porcine preprorelaxin and prorelaxin
CC   and synthetic A, B and C peptide chains of prolaxin, and a gene for
CC   expression of porcine preprorelaxin or prorelaxin, and their sub-
CC   units (see AAN30186).  They also claim a double-stranded DNA fragment
CC   for the expression of the signal peptide chain of porcine
CC   preprorelaxin comprising a coding strand and a complementary strand
CC   corresp. to a defined mRNA sequence (see AAN30187-N30194) which
CC   corresp. to the most homologous regions between the pig and rat cDNA
CC   sequences.  A probe (AAN30195) is also claimed.
XX
SQ   Sequence   182 AA;

Alignment Scores:
Pred. No.:              1.34e-08        Length:         182
Score:                  141.00          Matches:        41
Percent Similarity:     40.00%          Conservative:   7
Best Local Similarity:  34.17%          Mismatches:     32
Query Match:            22.20%          Indels:         40
DB:                     4               Gaps:           2

US-09-518-842-1_COPY_76_417 (1-342) x AAP30392 (1-182)

Qy      100 CTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCC 159
            ||||||:::|||   |||   |||   :::   :::|||      ||||||   :::
Db       63 LeuGluThrGlyProProAlaGluThrMetProSerSerIleThrLysAspAlaGluIle 82

Qy      160 TTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 219
            |||           ||||||:::|||||||||      |||||||||      ||||||
Db       83 LeuLysMetMetLeuGluPheValProAsnLeuProGlnGluLeuLysAlaThrLeuSer 102

Qy      220 GAAGGGCAGCCATCATTG---------------------------------------- 237
            |||   ||||||||||||
Db      103 GluArgGlnProSerLeuArgGluLeuGlnGlnSerAlaSerLysAspSerAsnLeuAsn 122

Qy      238 ------------AAGAAAATAATACTT------------------------------ 252
                        ||||||||||||||||
Db      123 PheGluGluPheLysLysIleIleLeuAsnArgGlnAsnGluAlaGluAspLysSerLeu 142

Qy      253 --------------------------------TCCCGCAAAAAGAGAAGTGGACGTCAC 279
                                             ||||||||||||||||      |||
Db      143 LeuGluLeuLysAsnLeuGlyLeuAspLysHisSerArgLysLysArgLeuPheArgMet 162

Qy      280 AGATTTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                        ||||||:::|||   |||                  :::||||||
Db      163 ThrLeuSerGluLysCysCysGlnValGlyCysIleArgLysAspIleAlaArgLeuCys 182

RESULT 10
AAR31958
ID   AAR31958 standard; Protein; 178 AA.
XX
AC   AAR31958;
XX
DT   17-DEC-2001  (updated)
DT   09-JUN-1993  (first entry)
XX
DE   Squamous cell specific SQ10 protein.
XX
KW   Squamous cell carcinoma; SCC; marker; differentiation;
KW   preprorelaxin; rabbit.
XX
OS   Oryctolagus cuniculus.
XX
PN   USN7783046-N.
XX
PD   15-DEC-1992.
XX
PF   28-OCT-1991;   91US-0783046.
XX
PR   28-OCT-1991;   91US-0783046.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Jetten AM,  Lotan R;
XX
DR   WPI; 1993-067425/08.
DR   N-PSDB; AAQ36775.
XX
PT   Diagnosis and assessment of squamous cell carcinoma - by
PT   identification of intracellular and secreted molecular markers of
PT   squamous cell differentiation
XX
PS   Example 1; Fig 1; 38pp; English.
XX
CC   A cDNA library was constructed from poly(A)+ RNA isolated from
CC   squamous differentiated rabbit tracheal epithelial cells.  The cDNA
CC   clone, SQ10, isolated from this library had a sequence homologous
CC   with those of the genes encoding human and porcine relaxin, i.e. the
CC   gene encodes a secreted protein related to preprorelaxin.  The
CC   protein is a marker for squamous cell differentiation, and is
CC   secreted extracellularly.  This allows body fluids e.g. sera, saliva
CC   and urine to be analysed in diagnosis of squamous cell carcinoma (SCC).
CC   Antibodies raised against the protein or immunogenic fragments
CC   react with squamous differentiated cells but not undifferentiated cells
CC   and are thus specific for a marker of squamous cell differentiation.
CC   See also AAR31959.
CC   (Note: Revised entry submitted to correct the patent number format of
CC   US Government-owned NTIS applications to prevent clashes with ongoing US
CC   granted patent numbers. For further information please visit the Derwent
CC   web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ   Sequence   178 AA;

Alignment Scores:
Pred. No.:              7.22e-05        Length:         178
Score:                  111.00          Matches:        39
Percent Similarity:     32.81%          Conservative:   3
Best Local Similarity:  30.47%          Mismatches:     42
Query Match:            17.48%          Indels:         44
DB:                     14              Gaps:           3

US-09-518-842-1_COPY_76_417 (1-342) x AAR31958 (1-178)

Qy       37 CACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGG 96
            ||||||      |||   ||||||      |||
Db       50 HisLeuGluArgGluSerProSerProGluAsnProPhe---------------------- 62

Qy       97 CTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAA 156
               |||   ||||||         |||   |||   :::|||      ||||||   :::
Db       63 ---LeuSerSerGlyProAlaAlaGluThrValProSerSerIleLysLysAspAlaGlu 81

Qy      157 GCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTG 216
                ||      |||   ||||||||||||      ||||||         |||
Db       82 AsnAlaAsnThrMetLeuGluSerIleProAsnLeuProGlnGluLeuThrAlaThrLeu 101

Qy      217 TCTGAAGGGCAGCCATCA------------------------------------------ 234
               |||   |||||||||
Db      102 PheGluLysGlnProSerLysLeuTyrLeuGlnTyrLeuProThrLeuLysLysSerAsn 121

Qy      235 ---------------TTGAAGAAAATAATA---------------------------- 249
                           ||||||||||||
Db      122 ValSerPheGluGluPheLysLysIleIleGlnAsnIleGlnArgGlyValGlnGlySer 141

Qy      250 --------------------CTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                 ||||||||||||||||
Db      142 SerAlaSerGluSerAsnThrPheSerArgLysLysArgGlnPheSerGluSerLeuPro 161
```

```
Qy        289 CCATTCTGTTGTGAAGTAATTTGT 312
                  ||||||:::      |||
Db        162 GluGluCysCysLysTyrGlyCys 169

RESULT 11
AAR64904
ID   AAR64904 standard; Protein; 150 AA.
XX
AC   AAR64904;
XX
DT   06-SEP-1995  (first entry)
XX
DE   Prolrelaxin fragment amino acids 12-161.
XX
KW   Prorelaxin; cervical ripening; ovarian peptide hormone;
KW   mammary gland development; skin elasticity; cardiovascular therapy;
KW   relaxin; premature labour; plasmid pTR21.
XX
OS   Synthetic.
XX
PN   WO9500645-A.
XX
PD   05-JAN-1995.
XX
PF   20-JUN-1994;   94WO-US06997.
XX
PR   21-JUN-1993;   93US-0080354.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Breece T,  Hayenga K,  Rinderknecht E,  Vandlen R;
PI   Yansura D;
XX
DR   WPI; 1995-052082/07.
DR   N-PSDB; AAQ76309.
XX
PT   Relaxin prodn. from non-natural recombinant prorelaxin - by
PT   cleavage of leader and C peptide chain, also new prorelaxin and
PT   related DNA, vectors etc.
XX
PS   Example 1; Fig 9A; 62pp; English.
XX
CC   AAQ76309 is the Not I-Bam HI restriction fragment of the plasmid
CC   pTR21, which encodes AAR64904, prorelaxin amino acids 12-161.
CC   Relaxin is produced by removing the non-naturally occuring leader
CC   and C-peptide from PR with a cleavage agent. Relaxin is an ovarian
CC   peptide hormone involved in the inhibition of premature labour,
CC   cervical ripening and the development of the mammary glands. It
CC   may also improve skin elasticity and has been used in cardiovascular
CC   therapy.
XX
SQ   Sequence   150 AA;

Alignment Scores:
Pred. No.:              0.000328        Length:         150
Score:                  105.50          Matches:        43
Percent Similarity:     35.67%          Conservative:   13
Best Local Similarity:  27.39%          Mismatches:     44
Query Match:            16.61%          Indels:         57
DB:                     16              Gaps:           6

US-09-518-842-1_COPY_76_417 (1-342) x AAR64904 (1-150)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db          7 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 26

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                ||:::              ||||||                          ||||||
Db         27 AlaProGln--------------ThrPro----------------------ArgPro 33

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                |||::::|||   :::      |||||||||   :::   :::         |||
Db         34 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 53

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||:::   ||||||      |||||||||      |||||||||   ||||||:::
Db         54 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 73

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db         74 LeuProGlnLeuGlnGlnHisValProValLeuLysAspSerSerLeuLeuPheGluGlu 93

Qy        238 ---AAGAAAATAATACTT---------------------------------------------- 252
                 ||||||:::|||
Db         94 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 113

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||||||||||||
Db        114 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 133

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                ||||||   |||   |||         :::      :::   |||
Db        134 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 150

RESULT 12
AAR07987
ID   AAR07987 standard; protein; 162 AA.
XX
AC   AAR07987;
XX
DT   27-FEB-1991  (first entry)
XX
DE   H2 prorelaxin gene encoded by insert in plasmid pTrpProRelAsp.
XX
KW   Relaxin; PR.
XX
OS   Homo sapiens.
XX
PN   WO9013659-A.
XX
PD   15-NOV-1990.
XX
PF   16-APR-1990;   90WO-US02085.
XX
PR   04-MAY-1989;   89US-0347550.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Henner DJ,  Vandlen RI,  Wikins JA,  Yansura D;
XX
DR   WPI; 1990-361494/48.
DR   N-PSDB; AAQ06595.
XX
PT   Cleaving polypeptide into polypeptide cleavage prods. - by
PT   treating free-cysteine form of polypeptide at desired junction.
XX
PS   Disclosure; Fig 2A; 56pp; English.
XX
CC   The plasmid was prepd. from a clone isolated from a cDNA library
CC   prepd. from RNA isolated from human corpus luteum, screened with an
CC   H1-cDNA probe.  A fragment encoding the N-terminal of PR was iso-
CC   lated and was ligated with a 410 bp fragment encoding AAs 17-153 of
CC   PR, and a fragment from pHGH2071*L (including the Trp promoter, AAs
CC   1-137 of metHGH, and amp and tet resistance.  A portion of the
CC   resulting construction (pFEproH2) was ligated with a fragment from
CC   the original clone encoding the PR C-terminal, and a fragment from
CC   pBR322XAP encoding part of the beta-lactamase gene. This produced
CC   plasmid pTrpProRel.  A StII signal sequence was then fused to the
CC   PR gene and a portion including the gene and signal sequence was
CC   ligated into a vector identical to pTrpStIIHGH in which the HGH
CC   gene had been removed.  The resulting construction, pTrpStIIProRel
CC   was treated to remove the StII sequence and the first 11 AAs of H2
CC   PR.  This was replaced with a synthetic DNA duplex encoding the
CC   first 12 AAs of H2 PR (including Asp1), to produce pTrpProRelAsp,
```

```
CC   the coding part of which is shown below.  The vector can be used to
CC   construct vectors that encode Asp-inserted H2PR with enhanced acid
CC   cleavage sites.
CC   See also AAR07988.
XX
SQ   Sequence   162 AA;

Alignment Scores:
Pred. No.:              0.000337      Length:         162
Score:                  105.50        Matches:        43
Percent Similarity:     35.67%        Conservative:   13
Best Local Similarity:  27.39%        Mismatches:     44
Query Match:            16.61%        Indels:         57
DB:                     11            Gaps:           6

US-09-518-842-1_COPY_76_417 (1-342) x AAR07987 (1-162)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         19 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 38

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::              ||||||                          ||||||
Db         39 AlaProGln--------------ThrPro----------------------ArgPro 45

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::      |||||||||   :::   :::         |||
Db         46 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 65

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||:::  ||||||      |||||||||      |||||||||   ||||||:::
Db         66 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 85

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db         86 LeuProGlnLeuGlnGlnHisValProValLeuLysAspSerSerLeuLeuPheGluGlu 105

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                 ||||||:::|||
Db        106 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 125

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||||||||||
Db        126 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 145

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                |||||   |||   |||         :::      :::   |||
Db        146 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 162

RESULT 13
AAP94621
ID   AAP94621 standard; protein; 185 AA.
XX
AC   AAP94621;
XX
DT   21-JUN-1990  (first entry)
XX
DE   Amino acid sequence of human preprorelaxin H2.
XX
KW   Relaxin; H2-relaxin; pubic symphysis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /label=Signal peptide
FT   Peptide         26..57
FT                   /label=B-chain
FT   Peptide         58..161
FT                   /label=C-peptide
FT   Peptide         162..159
FT                   /label=A-chain
XX
PN   EP303033-A.
XX
PD   15-FEB-1989.
XX
PF   12-DEC-1983;   83EP-0110103.
XX
PR   13-DEC-1982;   82AU-0007247.
XX
PA   (FLOR-) FLOREY H INST EXPER.
XX
PI   Hudson PJ,  Trgear GW,  Niall HD;
XX
DR   WPI; 1989-047874/07.
DR   N-PSDB; AAN92483.
XX
PT   New human H2-relaxin analogues -
PT   with shortened and/or modified A and/or B chains.
XX
PS   Disclosure; ; 25pp; English.
XX
CC   "H2" relaxin is deduced from a cDNA clone, has the general properties of
CC   a growth factor and is capable of altering nature of connective tissue
CC   and inducing smooth muscle contraction, specifically during labour.
XX
SQ   Sequence   185 AA;

Alignment Scores:
Pred. No.:              0.000354      Length:         185
Score:                  105.50        Matches:        43
Percent Similarity:     35.67%        Conservative:   13
Best Local Similarity:  27.39%        Mismatches:     44
Query Match:            16.61%        Indels:         57
DB:                     10            Gaps:           6

US-09-518-842-1_COPY_76_417 (1-342) x AAP94621 (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         42 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::              ||||||                          ||||||
Db         62 AlaProGln--------------ThrPro----------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::      |||||||||   :::   :::         |||
Db         69 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||:::  ||||||      |||||||||      |||||||||   ||||||:::
Db         89 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 108

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db        109 LeuProGlnLeuGlnGlnHisValProValLeuLysAspSerSerLeuLeuPheGluGlu 128

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                 ||||||:::|||
Db        129 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 148

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||||||||||
Db        149 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                |||||   |||   |||         :::      :::   |||
Db        169 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 185

RESULT 14
AAP40154
ID   AAP40154 standard; Protein; 185 AA.
XX
AC   AAP40154;
```

```
XX
DT   11-FEB-1992  (first entry)
XX
DE   Sequence of human preprorelaxin.
XX
KW   Labour; birth; hormone; relaxin.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /label= signal
FT   Peptide         26..57
FT                   /label= B-chain
FT   Peptide         58..161
FT                   /label= C-peptide
FT   Peptide         162..185
FT                   /label= A-chain
XX
PN   EP101309-A.
XX
PD   22-FEB-1984.
XX
PF   11-AUG-1983;   83EP-0304662.
XX
PR   12-AUG-1982;   82AU-0005352.
PR   11-AUG-1983;   83AU-0017906.
PR   01-JAN-1988;   88EP-0104503.
XX
PA   (FLOR-) HOWARD FLOREY INST.
PA   (HOWA-) HOWARD FLOREY INST.
XX
PI   Hudson PJ,  Shine J,  Niall HD,  Tregear GW;
XX
DR   WPI; 1984-050918/09.
DR   N-PSDB; AAN40101.
XX
PT   Genes for human relaxin, prorelaxin and preprorelaxin prodn. -
PT   prepd. by recombinant DNA techniques
XX
PS   Disclosure; Fig 2; 51pp; English.
XX
CC   The inventors claim the gene for the expression of human
CC   preprorelaxin (HPP) and its sub-units. Also claimed are synthetic
CC   HPP, HP, fragments and analogues. Human relaxin and its analogues
CC   are prepd. for therapeutic purposes, esp. in clinical intervention
CC   in cases of difficult labour.
XX
SQ   Sequence   185 AA;

Alignment Scores:
Pred. No.:              0.000628      Length:          185
Score:                  103.50        Matches:         44
Percent Similarity:     35.67%        Conservative:    12
Best Local Similarity:  28.03%        Mismatches:      44
Query Match:            16.30%        Indels:          57
DB:                     5             Gaps:            6

US-09-518-842-1_COPY_76_417 (1-342) x AAP40154 (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::     ||||||        :::   |||      ||||||
Db         42 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||::              ||||||                          ||||||
Db         62 AlaProGln--------------ThrPro----------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                 |||:::|||   :::      |||||||||   :::   :::
Db         69 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeu 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
                 |||||||||   ||||||   ||||||||||||      |||||||||   |||||||||
Db         89 GluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSerGluArgGlnProSer 108

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db        109 LeuProGluLeuGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGlu 128

Qy        238 ---AAGAAAATAATACTTTCCCGC---------------------------------- 258
                 ||||||:::|||   :::|||
Db        129 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeu 148

Qy        259 ----------------------------AAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                           |||||||||
Db        149 LysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPhe 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                 ||||||   :::   |||        :::      |||   |||
Db        169 GluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaLysTyrCys 185

RESULT 15
AAP40155
ID   AAP40155 standard; Protein; 185 AA.
XX
AC   AAP40155;
XX
DT   11-FEB-1992  (first entry)
XX
DE   Sequence of human preprorelaxin.
XX
KW   Labour; birth; hormone; relaxin.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /label= signal
FT   Peptide         26..57
FT                   /label= B-chain
FT   Peptide         58..161
FT                   /label= C-peptide
FT   Peptide         162..185
FT                   /label= A-chain
XX
PN   EP101309-A.
XX
PD   22-FEB-1984.
XX
PF   11-AUG-1983;   83EP-0304662.
XX
PR   12-AUG-1982;   82AU-0005352.
PR   11-AUG-1983;   83AU-0017906.
PR   01-JAN-1988;   88EP-0104503.
XX
PA   (FLOR-) HOWARD FLOREY INST.
PA   (HOWA-) HOWARD FLOREY INST.
XX
PI   Hudson PJ,  Shine J,  Niall HD,  Tregear GW;
XX
DR   WPI; 1984-050918/09.
DR   N-PSDB; AAN40123, AAN40124.
XX
PT   Genes for human relaxin, prorelaxin and preprorelaxin prodn. -
PT   prepd. by recombinant DNA techniques
XX
PS   Disclosure; Fig 3; 51pp; English.
XX
CC   The inventors claim the gene for the expression of human
CC   preprorelaxin (HPP) and its sub-units. Also claimed are synthetic
CC   HPP, HP, fragments and analogues. Human relaxin and its analogues
CC   are prepd. for therapeutic purposes, esp. in clinical intervention
CC   in cases of difficult labour.
XX
SQ   Sequence   185 AA;
```

Alignment Scores:

Pred. No.:	0.000628	Length:	185
Score:	103.50	Matches:	44
Percent Similarity:	35.67%	Conservative:	12
Best Local Similarity:	28.03%	Mismatches:	44
Query Match:	16.30%	Indels:	57
DB:	5	Gaps:	6

US-09-518-842-1_COPY_76_417 (1-342) x AAP40155 (1-185)

```
Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         42 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                |||:::              ||||||                      ||||||
Db         62 AlaProGln--------------ThrPro----------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                   |||:::|||   :::      |||||||||   :::   :::
Db         69 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeu 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||||   ||||||   ||||||||||||      |||||||||   |||||||||
Db         89 GluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSerGluArgGlnProSer 108

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db        109 LeuProGluLeuGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGlu 128

Qy        238 ---AAGAAAATAATACTTTCCCGC---------------------------------- 258
                 ||||||:::|||   :::|||
Db        129 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeu 148

Qy        259 ----------------------------AAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                         |||||||||
Db        149 LysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPhe 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                 ||||||   :::   |||         :::      |||   |||
Db        169 GluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaLysTyrCys 185
```

Search completed: June 14, 2003, 19:44:47
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:37:50 ; Search time 19 Seconds
(without alignments)
3460.839 Million cell updates/sec

Title: US-09-518-842-1_COPY_76_417
Perfect score: 635
Sequence: 1 GCAGAGCTGAGGGGATGTGG.....CTTCAGTTAAATTATGTACA 342

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09518842/runat_14062003_175811_12284/app_query.fasta_1.519
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09518842_@CGN_1_1_62_@runat_14062003_175811_12284 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	141	22.2	182	1	RXPG	relaxin precursor
	2	116.5	18.3	185	2	S48082	relaxin precursor
	3	111	17.5	178	2	A49014	20K protein - rabb
	4	105.5	16.6	166	2	S42786	relaxin 2 precurso
	5	105.5	16.6	185	1	A60982	relaxin 2 precurso
	6	103.5	16.3	143	2	I47053	relaxin B,C and A
	7	103.5	16.3	185	1	A44559	relaxin 1 precurso
	8	103.5	16.3	185	2	A34936	relaxin precursor
	9	101	15.9	560	2	T16833	hypothetical prote
	10	100.5	15.8	166	2	S42783	relaxin 1 precurso
	11	97.5	15.4	186	1	RXRT	relaxin precursor
	12	83	13.1	321	2	E83259	hypothetical prote
c	13	78	13.0	754	2	AE0614	probable competenc
	14	77	12.1	457	2	AG0433	proteinase (EC 3.4
	15	76.5	12.0	160	2	A49194	relaxin - guinea p
	16	74.5	11.7	1531	2	H71468	probable outer mem
	17	73.5	11.6	273	2	G89979	NAD synthetase, pr
c	18	73.5	12.2	402	2	A84581	probable disease r
c	19	73	12.2	993	2	S46779	26S proteasome reg
	20	73	11.5	1435	2	S69632	regulatory protein
	21	72.5	11.4	292	2	AB0825	probable membrane
	22	72.5	11.4	1259	2	T16038	hypothetical prote
c	23	72.5	12.1	1744	1	C4HU	complement C4A pre
	24	72	11.3	391	2	S60672	replication protei
	25	72	11.3	453	2	I50674	retinoic acid rece
	26	72	11.3	483	2	AD3618	succinate-semialde
	27	72	11.3	752	2	T50450	hypothetical prote
	28	72	11.3	836	1	JDVLD	DNA-directed DNA p
	29	72	11.3	2397	1	A55535	versican precursor
c	30	71.5	11.9	1217	2	T25894	hypothetical prote
c	31	71	11.8	206	2	S61705	hypothetical prote
	32	71	11.2	1490	2	S72351	nonstructural poly
	33	70.5	11.1	271	1	Z8BPC2	gene 18 protein -
	34	70.5	11.1	347	2	T41373	hypothetical prote
c	35	70.5	11.8	766	2	A86198	hypothetical prote
	36	70.5	11.1	957	2	F69729	excinuclease ABC c
	37	70	11.0	283	2	S55486	probable pantoate-
	38	69.5	10.9	479	2	B87092	GTP1/Obg-family GT
	39	69.5	10.9	645	2	E84042	threonyl-tRNA synt
	40	69.5	10.9	1571	2	AC1647	Lactobacillus phag
c	41	69	11.5	274	1	MNVNVM	nonstructural prot
c	42	69	11.5	349	2	T25150	hypothetical prote
	43	69	10.9	508	2	H95271	probable long chai
c	44	69	11.5	544	2	AB0716	probable substrate
c	45	69	11.5	570	2	T37314	probable kexin (EC

ALIGNMENTS

RESULT 1
RXPG
relaxin precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 18-Jun-1999
C;Accession: A90934; A93187; A90205; A90196; S32313; S32312; A29796; A01615
R;Haley, J.; Hudson, P.; Scanlon, D.; John, M.; Cronk, M.; Shine, J.; Tregear, G.; Ni
DNA 1, 155-162, 1982
A;Title: Porcine relaxin: molecular cloning and cDNA structure.
A;Reference number: A90934; MUID:83157118; PMID:6897721
A;Accession: A90934
A;Molecule type: DNA
A;Residues: 1-182 <HAL>
A;Cross-references: GB:K01088; NID:g164634; PIDN:AAA31114.1; PID:g164635
R;James, R.; Niall, H.; Kwok, S.; Bryant-Greenwood, G.
Nature 267, 544-546, 1977
A;Title: Primary structure of porcine relaxin: homology with insulin and related grow
A;Reference number: A93187; MUID:77213067; PMID:876374
A;Accession: A93187
A;Molecule type: protein
A;Residues: 25-50,'TWGR';161-182 <JAM>
R;Schwabe, C.; McDonald, J.K.; Steinetz, B.G.
Biochem. Biophys. Res. Commun. 75, 503-510, 1977
A;Title: Primary structure of the B-chain of porcine relaxin.
A;Reference number: A90205; MUID:77157271; PMID:851452
A;Accession: A90205
A;Molecule type: protein
A;Residues: 25-47,'VW',50 <SCH1>
R;Schwabe, C.; McDonald, J.K.
Biochem. Biophys. Res. Commun. 74, 1501-1504, 1977
A;Title: Demonstration of a pyroglutamyl residue at the N terminus of the B-chain of
A;Reference number: A90201; MUID:77134136; PMID:843375
A;Contents: annotation; pyrrolidone carboxylic acid
R;Schwabe, C.; McDonald, J.K.; Steinetz, B.G.
Biochem. Biophys. Res. Commun. 70, 397-405, 1976
A;Title: Primary structure of the A chain of porcine relaxin.
A;Reference number: A90196; MUID:76231539; PMID:938497
A;Accession: A90196

A;Molecule type: protein
A;Residues: 161-169,'E',171-182 <SCH>
R;Schwabe, C.; McDonald, J.K.
Science 197, 914-915, 1977
A;Title: Relaxin: a disulfide homolog of insulin.
A;Reference number: A94245; MUID:77236040; PMID:887933
A;Contents: annotation; disulfide bonds
R;Kohsaka, T.; Takahara, H.; Sugawara, K.; Tagami, S.
Biol. Chem. Hoppe-Seyler 374, 203-210, 1993
A;Title: Endogenous heterogeneity of relaxin and sequence of the major form in pregnant
A;Reference number: S32312; MUID:93257096; PMID:8489740
A;Accession: S32313
A;Molecule type: protein
A;Residues: 25-53 <KOH>
A;Accession: S32312
A;Molecule type: protein
A;Residues: 161-182 <KO2>
R;Haley, J.; Crawford, R.; Hudson, P.; Scanlon, D.; Tregear, G.; Shine, J.; Niall, H.
J. Biol. Chem. 262, 11940-11946, 1987
A;Title: Porcine relaxin. Gene structure and expression.
A;Reference number: A29796; MUID:87308187; PMID:2442155
A;Accession: A29796
A;Molecule type: DNA
A;Residues: 1-115,'L',117-182 <HA2>
A;Cross-references: GB:J02792; NID:g164636; PIDN:AAA31115.1; PID:g164637
C;Comment: Relaxin is an ovarian hormone that acts with estrogen to produce dilation of
C;Genetics:
A;Introns: 70/1
C;Superfamily: insulin
C;Keywords: pyroglutamic acid
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-56/Domain: relaxin chain B #status experimental <RCB>
F;25-56,161-182/Product: relaxin #status experimental <MAT>
F;161-182/Domain: relaxin chain A #status experimental <RCA>
F;25/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen
F;34-169,46-182,168-173/Disulfide bonds: #status experimental

Alignment Scores:

Pred. No.:	2.74e-07	Length:	182
Score:	141.00	Matches:	41
Percent Similarity:	40.00%	Conservative:	7
Best Local Similarity:	34.17%	Mismatches:	32
Query Match:	22.20%	Indels:	40
DB:	1	Gaps:	2

US-09-518-842-1_COPY_76_417 (1-342) x RXPG (1-182)

```
Qy        100 CTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCC 159
              |||||||:::|||   |||   |||   :::   :::|||     ||||||   :::
Db         63 LeuGluThrGlyProProAlaGluThrMetProSerSerIleThrLysAspAlaGluIle 82

Qy        160 TTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 219
              |||           ||||||:::||||||||||     |||||||||     ||||||
Db         83 LeuLysMetMetLeuGluPheValProAsnLeuProGlnGluLeuLysAlaThrLeuSer 102

Qy        220 GAAGGGCAGCCATCATTG---------------------------------------- 237
              |||   ||||||||||||
Db        103 GluArgGlnProSerLeuArgGluLeuGlnGlnSerAlaSerLysAspSerAsnLeuAsn 122

Qy        238 ------------AAGAAAATAATACTT------------------------------ 252
                         ||||||||||||||
Db        123 PheGluGluPheLysLysIleIleLeuAsnArgGlnAsnGluAlaGluAspLysSerLeu 142

Qy        253 ---------------------------------TCCCGCAAAAAGAGAAGTGGACGTCAC 279
                                              ||||||||||||||     |||
Db        143 LeuGluLeuLysAsnLeuGlyLeuAspLysHisSerArgLysLysArgLeuPheArgMet 162

Qy        280 AGATTTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                          ||||||:::|||   |||                   :::||||||
Db        163 ThrLeuSerGluLysCysCysGlnValGlyCysIleArgLysAspIleAlaArgLeuCys 182
```

RESULT 2
S48082

relaxin precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C;Accession: S48082; PC2067; PN0626
R;Evans, B.A.; John, M.; Fowler, K.J.; Summers, R.J.; Cronk, M.; Shine, J.; Tregear,
J. Mol. Endocrinol. 10, 15-23, 1993
A;Title: The mouse relaxin gene: nucleotide sequence and expression.
A;Reference number: S48082; MUID:93199663; PMID:8452637
A;Accession: S48082
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-185 <EVA>
A;Cross-references: EMBL:Z27088; NID:g414780; PIDN:CAA81611.1; PID:g414781
R;Buellesbach, E.E.; Schwabe, C.
Biochem. Biophys. Res. Commun. 196, 311-319, 1993
A;Title: Mouse relaxin: synthesis and biological activity of the first relaxin with a
A;Reference number: PN0626; MUID:94030011; PMID:8216305
A;Accession: PC2067
A;Molecule type: protein
A;Residues: 23-57 <BUE>
A;Accession: PN0626
A;Molecule type: protein
A;Residues: 161-185 <BU2>
A;Note: proteins with and without 184-Tyr were synthesized, their biological activiti
C;Superfamily: insulin
C;Keywords: hormone
F;23-57,161-185/Product: relaxin #status experimental <MAT>
F;23-57/Domain: chain B #status experimental <CHB>
F;161-185/Domain: chain A #status experimental <CHA>
F;36-171,48-185,170-175/Disulfide bonds: #status experimental

Alignment Scores:

Pred. No.:	0.000147	Length:	185
Score:	116.50	Matches:	38
Percent Similarity:	35.76%	Conservative:	16
Best Local Similarity:	25.17%	Mismatches:	56
Query Match:	18.35%	Indels:	41
DB:	2	Gaps:	2

US-09-518-842-1_COPY_76_417 (1-342) x S48082 (1-185)

```
Qy          7 CTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAG 66
              :::|||   ||||||      :::   :::   |||:::      |||
Db         33 IleArgMetCysGlyArgGluTyrAlaArgGluLeuIleLysIleCysGlyAlaSerVal 52

Qy         67 AAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATG 126
                          :::                |||   :::   :::      |||:::
Db         53 GlyArgLeuAlaLeuSerGlnGluGluProAlaLeuLeuAlaArgGlnAlaThrGluVal 72

Qy        127 GTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCT 186
              |||   :::       |||||||||   :::         ||||||   :::   :::|||
Db         73 ValProSerPheIleAsnLysAspAlaGluProPheAspThrThrLeuLysCysLeuPro 92

Qy        187 AATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATA 246
              |||||||||   |||||||||      |||||||||   |||   ||||||   ::::::
Db         93 AsnLeuSerGluGluLeuLysAlaValLeuSerGluAlaGlnAlaSerLeuProGluLeu 112

Qy        247 ------------ATACTT------------------------------------------ 252
                         :::|||
Db        113 GlnHisAlaProValLeuSerAspSerValValSerLeuGluGlyPheLysLysThrLeu 132

Qy        252 ------------------------------------------------------------ 252

Db        133 HisAspArgLeuGlyGluAlaGluAspGlySerProProGlyLeuLysTyrLeuGlnSer 152

Qy        253 ---------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAA 303
                       ||||||||||||||                                ||||||
Db        153 AspThrHisSerArgLysLysArgGluSerGlyGlyLeuMetSerGlnGlnCysCysHis 172

Qy        304 GTAATTTGTGACGATGGAACTTCAGTTAAATTA 336
              |||   |||         :::      ||||||
Db        173 ValGlyCysSerArgArgSerIleAlaLysLeu 183
```

```
RESULT 3
A49014
20K protein - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C;Accession: A49014
R;Jetten, A.M.; Bernacki, S.H.; Floyd, E.E.; Saunders, N.A.; Pieniazek, J.; Lotan, R.
Cell Growth Differ. 3, 549-556, 1992
A;Title: Expression of a preprorelaxin-like gene during squamous differentiation of rabb
A;Reference number: A49014; MUID:93002619; PMID:1339318
A;Accession: A49014
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-178 <JET>
A;Cross-references: GB:S45940; NID:g257388; PIDN:AAB23648.1; PID:g257389
A;Experimental source: tracheobronchial epithelial cells
A;Note: sequence extracted from NCBI backbone (NCBIN:115816, NCBIP:115821)
C;Superfamily: insulin

Alignment Scores:
Pred. No.:              0.0006        Length:         178
Score:                  111.00        Matches:        39
Percent Similarity:     32.81%        Conservative:   3
Best Local Similarity:  30.47%        Mismatches:     42
Query Match:            17.48%        Indels:         44
DB:                     2             Gaps:           3

US-09-518-842-1_COPY_76_417 (1-342) x A49014 (1-178)

Qy         37 CACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGG 96
              ||||||           |||   ||||||      |||
Db         50 HisLeuGluArgGluSerProSerProGluAsnProPhe---------------------- 62

Qy         97 CTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAA 156
                 |||   ||||||         |||   |||   :::|||      ||||||   :::
Db         63 ---LeuSerSerGlyProAlaAlaGluThrValProSerSerIleLysLysAspAlaGlu 81

Qy        157 GCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTG 216
                    |||      |||   ||||||||||||      ||||||         |||
Db         82 AsnAlaAsnThrMetLeuGluSerIleProAsnLeuProGlnGluLeuThrAlaThrLeu 101

Qy        217 TCTGAAGGGCAGCCATCA---------------------------------------- 234
                 |||   |||||||||
Db        102 PheGluLysGlnProSerLysLeuTyrLeuGlnTyrLeuProThrLeuLysLysSerAsn 121

Qy        235 ---------------TTGAAGAAAATAATA---------------------------- 249
                            ||||||||||||
Db        122 ValSerPheGluGluPheLysLysIleIleGlnAsnIleGlnArgGlyValGlnGlySer 141

Qy        250 --------------------CTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                   ||||||||||||||||
Db        142 SerAlaSerGluSerAsnThrPheSerArgLysLysArgGlnPheSerGluSerLeuPro 161

Qy        289 CCATTCTGTTGTGAAGTAATTTGT 312
                  ||||||:::     |||
Db        162 GluGluCysCysLysTyrGlyCys 169

RESULT 4
S42786
relaxin 2 precursor - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C;Accession: S42786
R;Evans, B.B.
submitted to the EMBL Data Library, November 1993
A;Reference number: S42776
A;Accession: S42786
A;Molecule type: mRNA
A;Residues: 1-166 <EVA>
A;Cross-references: EMBL:Z27245; NID:g416109; PIDN:CAA81758.1; PID:g416110
C;Genetics:
A;Gene: rlx2
C;Superfamily: insulin
F;1-5/Domain: signal sequence (fragment) #status predicted <SIG>
F;6-166/Product: relaxin 2 #status predicted <MAT>

Alignment Scores:
Pred. No.:              0.00244       Length:         166
Score:                  105.50        Matches:        43
Percent Similarity:     35.67%        Conservative:   13
Best Local Similarity:  27.39%        Mismatches:     44
Query Match:            16.61%        Indels:         57
DB:                     2             Gaps:           6

US-09-518-842-1_COPY_76_417 (1-342) x S42786 (1-166)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         23 AlaGlnIleAlaIleCysGlyLysSerThrTrpSerLysArgSerLeuSerGlnGluAsp 42

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::               ||||||                          ||||||
Db         43 AlaProGln--------------ThrPro-----------------------ArgPro 49

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                    |||:::|||   :::      |||||||||   :::   :::          |||
Db         50 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 69

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
                 ||||||:::   ||||||      |||||||||      |||||||||   ||||||:::
Db         70 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 89

Qy        235 TTG---------------------------------------------------------- 237
                 |||
Db         90 LeuProGlnLeuGlnGlnTyrValProValLeuLysAspSerSerLeuLeuPheGluGlu 109

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                    ||||||:::|||
Db        110 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 129

Qy        253 ------------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                       ||||||||||||||||
Db        130 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 149

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                  ||||||   |||   |||         :::      :::   |||
Db        150 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 166

RESULT 5
A60982
relaxin 2 precursor [validated] - human
N;Alternate names: preprorelaxin 2
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 08-Dec-2000
C;Accession: A05092; A60982
R;Hudson, P.; John, M.; Crawford, R.; Haralambidis, J.; Scanlon, D.; Gorman, J.; Treg
EMBO J. 3, 2333-2339, 1984
A;Title: Relaxin gene expression in human ovaries and the predicted structure of a hu
A;Reference number: A05092; MUID:85051298; PMID:6548702
A;Accession: A05092
A;Molecule type: mRNA
A;Residues: 1-185 <HUD>
A;Cross-references: GB:X00948; NID:g35926; PIDN:CAA25460.1; PID:g35927
R;Stults, J.T.; Bourell, J.H.; Canova-Davis, E.; Ling, V.T.; Laramee, G.R.; Winslow,
Biomed. Environ. Mass Spectrom. 19, 655-664, 1990
A;Title: Structural characterization by mass spectrometry of native and recombinant h
A;Reference number: A60982; MUID:91167739; PMID:2076464
A;Accession: A60982
A;Molecule type: protein
A;Residues: 25-53;162-185 <STU>
C;Genetics:
A;Gene: GDB:RLN2
A;Cross-references: GDB:119553; OMIM:179740
A;Map position: 9pter-9q12
C;Superfamily: insulin
C;Keywords: ovary; pyroglutamic acid
```

```
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-53/Domain: relaxin 2 chain B #status experimental <BCH>
F;25-53,162-185/Product: relaxin 2 #status experimental <MAT>
F;58-157/Domain: relaxin 2 connecting C peptide #status predicted <CPEP>
F;162-185/Domain: relaxin 2 chain A #status experimental <ACH>
F;35-172,47-185,171-176/Disulfide bonds: #status experimental
F;162/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Alignment Scores:
Pred. No.:                0.00247       Length:          185
Score:                    105.50        Matches:         43
Percent Similarity:       35.67%        Conservative:    13
Best Local Similarity:    27.39%        Mismatches:      44
Query Match:              16.61%        Indels:          57
DB:                       1             Gaps:            6

US-09-518-842-1_COPY_76_417 (1-342) x A60982 (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||:::::::      ||||||         :::   |||      ||||||
Db         42 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::               ||||||                          ||||||
Db         62 AlaProGln--------------ThrPro-----------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                  |||:::|||   :::      |||||||||   :::   :::         |||
Db         69 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleAsnMetMetSer 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
                 ||||||:::   ||||||      |||||||||      |||||||||   ||||||:::
Db         89 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluMetGlnProAla 108

Qy        235 TTG------------------------------------------------------- 237
                 |||
Db        109 LeuProGlnLeuGlnGlnHisValProValLeuLysAspSerSerLeuLeuPheGluGlu 128

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                 ||||||:::|||
Db        129 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerSerProSerGluLeu 148

Qy        253 ----------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                     ||||||||||||||||
Db        149 LysTyrLeuGlyLeuAspThrHisSerArgLysLysArgGlnLeuTyrSerAlaLeuAla 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                  ||||||   |||   |||         :::      :::   |||
Db        169 AsnLysCysCysHisValGlyCysThrLysArgSerLeuAlaArgPheCys 185

RESULT 6
I47053
relaxin B,C and A chains - horse (fragment)
C;Species: Equus sp.
C;Date: 04-Sep-1997 #sequence_revision 07-Nov-1997 #text_change 16-Jul-1999
C;Accession: I47053
R;Klonisch, T.; Ryan, P.L.; Yamashiro, S.; Porter, D.G.
Biol. Reprod. 52, 1307-1315, 1995
A;Title: Partial complementary deoxyribonucleic acid cloning of equine relaxin messenger
A;Reference number: I47053; MUID:95359320; PMID:7543295
A;Accession: I47053
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-143 <KLO>
A;Cross-references: GB:S78800; NID:g1042059; PIDN:AAB35036.1; PID:g1042060
C;Superfamily: insulin

Alignment Scores:
Pred. No.:                0.00403       Length:          143
Score:                    103.50        Matches:         30
Percent Similarity:       51.76%        Conservative:    14
Best Local Similarity:    35.29%        Mismatches:      34
Query Match:              16.30%        Indels:          7
DB:                       2             Gaps:            3

US-09-518-842-1_COPY_76_417 (1-342) x I47053 (1-143)

Qy          7 CTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCT--- 63
              ::::::   ||||||          :::      :::      |||
Db          1 IleLysAlaCysGlyArgGluLeuAlaArgLeuArgIleGluIleCysGlySerLeuSer 20

Qy         64 ---GAGAAGACATTC------ACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGT 114
                 :::||||||                 ||||||         ||||||   |||:::
Db         21 TrpLysLysThrValLeuArgLeuGluGluProGly---------LeuGluValGlyGln 37

Qy        115 CCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                 |||   |||:::|||||||:::|||   :::|||||   :::||||||   |||
Db         38 ProValGluIleValSerSerSerIleSerLysAspAlaGluAlaLeuAsnThrLysLeu 57

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
                        ||||||      |||   |||      |||||||||   |||||||||
Db         58 GlyLeuAsnSerAsnLeuProLysGluGlnLysAlaThrLeuSerGluArgGlnProSer 77

Qy        235 TTGAAGAAAATAATA 249
                 ::::::::::::
Db         78 TrpArgGluLeuLeu 82

RESULT 7
A44559
relaxin 1 precursor - human
N;Alternate names: preprorelaxin 1
C;Species: Homo sapiens (man)
C;Date: 03-Feb-1994 #sequence_revision 07-Oct-1994 #text_change 18-Jun-1999
C;Accession: B05092; A44559
R;Hudson, P.; John, M.; Crawford, R.; Haralambidis, J.; Scanlon, D.; Gorman, J.; Treg
EMBO J. 3, 2333-2339, 1984
A;Title: Relaxin gene expression in human ovaries and the predicted structure of a hu
A;Reference number: A05092; MUID:85051298; PMID:6548702
A;Accession: B05092
A;Molecule type: DNA
A;Residues: 1-185 <HU1>
A;Cross-references: GB:X00949; NID:g35932; PIDN:CAA25461.1; PID:g35933
R;Hudson, P.; Haley, J.; John, M.; Cronk, M.; Crawford, R.; Haralambidis, J.; Tregear
Nature 301, 628-631, 1983
A;Title: Structure of a genomic clone encoding biologically active human relaxin.
A;Reference number: A44559; MUID:83141755; PMID:6298628
A;Accession: A44559
A;Molecule type: DNA
A;Residues: 1-185 <HU2>
A;Cross-references: GB:X00949; NID:g35932; PIDN:CAA25461.1; PID:g35933
C;Comment: Relaxin is an ovarian hormone that acts with estrogen to produce dilation
 disulfide bonds.
C;Genetics:
A;Gene: GDB:RLN1
A;Cross-references: GDB:119552; OMIM:179730
A;Map position: 9pter-9q12
C;Superfamily: insulin
C;Keywords: hormone; ovary
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-57/Domain: relaxin 1 chain B #status predicted <BCH>
F;26-57,162-185/Product: relaxin 1 #status predicted <MAT>
F;58-158/Domain: relaxin 1 connecting C peptide #status predicted <CPEP>
F;162-185/Domain: relaxin 1 chain A #status predicted <ACH>
F;35-172,47-185,171-176/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.:                0.00412       Length:          185
Score:                    103.50        Matches:         44
Percent Similarity:       35.67%        Conservative:    12
Best Local Similarity:    28.03%        Mismatches:      44
Query Match:              16.30%        Indels:          57
DB:                       1             Gaps:            6

US-09-518-842-1_COPY_76_417 (1-342) x A44559 (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
```

```
                ||||:::::::      ||||||         :::   |||      ||||||
Db         42 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                ||||::          ||||||                              ||||||
Db         62 AlaProGln--------------ThrPro-----------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                 |||:::|||   :::      |||||||||   :::   :::
Db         69 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeu 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
                |||||||||   ||||||   ||||||||||||      |||||||||   |||||||||
Db         89 GluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSerGluArgGlnProSer 108

Qy        235 TTG---------------------------------------------------------- 237
                |||
Db        109 LeuProGluLeuGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGlu 128

Qy        238 ---AAGAAAATAATACTTTCCCGC---------------------------------- 258
                  |||||||:::|||   :::|||
Db        129 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeu 148

Qy        259 ----------------------------AAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                           |||||||||
Db        149 LysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgArgProTyrValAlaLeuPhe 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                 ||||||   :::   |||      :::      |||   |||
Db        169 GluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaLysTyrCys 185

RESULT 8
A34936
relaxin precursor - rhesus macaque
C;Species: Macaca mulatta (rhesus macaque)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 12-Jun-1998
C;Accession: A34936
R;Crawford, R.J.; Hammond, V.E.; Roche, P.J.; Johnston, P.D.; Tregear, G.W.
J. Mol. Endocrinol. 3, 169-174, 1989
A;Title: Structure of rhesus monkey relaxin predicted by analysis of the single-copy rhe
A;Reference number: A34936; MUID:90073957; PMID:2590381
A;Accession: A34936
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-185 <CRA>
C;Superfamily: insulin
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-185/Product: relaxin #status predicted <MAT>

Alignment Scores:
Pred. No.:              0.00412       Length:          185
Score:                  103.50        Matches:          41
Percent Similarity:     34.39%        Conservative:     13
Best Local Similarity:  26.11%        Mismatches:       46
Query Match:            16.30%        Indels:           57
DB:                     2             Gaps:              5

US-09-518-842-1_COPY_76_417 (1-342) x A34936 (1-185)

Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         ||||||      |||:::
Db         42 AlaGlnIleAlaIleCysGlyLysSerThrLeuGlyLysArgSerLeuAsnGlnGluAsp 61

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                |||   |||                                             ||||||
Db         62 AlaProLeuLysPro-----------------------------------------ArgPro 68

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                 |||:::|||   :::      |||:::|||   :::   :::          |||
Db         69 AlaAlaGluIleValProSerLeuIleAsnGlnAspThrGluThrIleAsnMetMetSer 88

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              ||||||:::   ||||||      |||||||||      |||||||||   |||||||:::
```

```
Db         89 GluPheValAlaAsnLeuProGlnGluLeuLysLeuThrLeuSerGluArgGlnProAla 108

Qy        235 TTG---------------------------------------------------------- 237
                |||
Db        109 LeuSerGluLeuGlnGlnHisValProValLeuLysAspSerAsnLeuSerPheGluGlu 128

Qy        238 ---AAGAAAATAATACTT---------------------------------------- 252
                  ||||||||||||
Db        129 PheLysLysIleIleArgLysArgGlnSerGluAlaThrAspSerSerProSerGluLeu 148

Qy        253 ------------------------TCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGAT 288
                                       |||||||:::||||||
Db        149 ArgSerLeuGlyLeuAspThrHisSerArgArgLysArgGlnLeuTyrMetThrLeuSer 168

Qy        289 CCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                 ||||||   :::   |||      :::      |||   |||
Db        169 AsnLysCysCysHisIleGlyCysThrLysLysSerLeuAlaLysPheCys 185

RESULT 9
T16833
hypothetical protein T07H6.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C;Accession: T16833
R;Geisel, C.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid T07H6.
A;Reference number: Z18586
A;Accession: T16833
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-560 <GEI>
A;Cross-references: EMBL:U53344; NID:g1255886; PID:g1255889; PIDN:AAA96225.1; GSPDB:G
A;Experimental source: strain Bristol N2; clone T07H6
C;Genetics:
A;Gene: CESP:T07H6.5
A;Map position: X
A;Introns: 14/1; 75/1; 102/3; 128/1; 186/1; 272/2; 326/1; 361/1; 422/1; 475/1; 527/1
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology

Alignment Scores:
Pred. No.:              0.00856       Length:          560
Score:                  101.00        Matches:          43
Percent Similarity:     38.56%        Conservative:     16
Best Local Similarity:  28.10%        Mismatches:       38
Query Match:            15.91%        Indels:           56
DB:                     2             Gaps:              9

US-09-518-842-1_COPY_76_417 (1-342) x T16833 (1-560)

Qy         19 GGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACC 78
              ||||||          |||   |||   |||:::|||      |||   ||||||:::
Db        414 GlyProProMet---LysCysLeuProSerTrpCysGluHisProSerLysThrTyrGly 432

Qy         79 ACCACCCCAGGAGGGTGGCTGCTG---------------------------------- 102
                |||   |||||||||   :::|||
Db        433 ThrLeuProGlyGlyGlnIleLeuLeuGluGlyIleLeuGlyAlaTyrGluPheGlnSer 452

Qy        103 --------------GAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 147
                            |||   ||||||            :::|||            |||
Db        453 TyrIleGlnLysValGluGluGlyArg---------AlaIleSerPheGlnCysGlyLys 469

Qy        148 GATGGACAAGCCTTAGGT-----------------ACGACATCAGAATTCATTCCTAAT 189
                          :::|||                           |||:::::::|||
Db        470 GlyAsnTyrLeuIleGlyProProLysAlaThrCysValAsnGlyGluTrpMetProLys 489

Qy        190 TTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG---AAAATA 246
              :::||||||         |||   :::|||:::      |||   ::::::   ||||||
Db        490 ValSerPro---------LysCysValSerGlnThrHisProMetIleGluGlyLysIle 506

Qy        247 ATACTTTCCCGCAAAAAGAGAAGT------GGACGT---------------------- 276
              :::      ||||||||||||||||      ||||||
Db        507 LeuTrpAspArgLysLysArgSerLeuProGlyArgAlaValArgGluTyrValAspAsp 526
```

```
Qy        277 ------------CACAGATTTGATCCATTCTGTTGTGAAGTAATT--------------- 309
                          ||||||            |||   ::::::
Db        527 GluLeuSerThrHisArgGlnHisSerGlyLysCysGlyIleValSerGlyLysLeuGlu 546

Qy        310 ------------------TGTGACGATGGAACTTCAGTT 330
                               |||:::|||   ||||||
Db        547 ArgMetIleMetGlnHisSerAspAsnGlyValSerVal 559
```

RESULT 10
S42783
relaxin 1 precursor - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C;Accession: S42783
R;Evans, B.B.
submitted to the EMBL Data Library, November 1993
A;Reference number: S42776
A;Accession: S42783
A;Molecule type: mRNA
A;Residues: 1-166 <EVA>
A;Cross-references: EMBL:Z27225; NID:g415996; PIDN:CAA81739.1; PID:g415997
C;Genetics:
A;Gene: rlx1
C;Superfamily: insulin
C;Keywords: disulfide bond; hormone
F;1-5/Domain: signal sequence (fragment) #status predicted <SIG>
F;6-166/Product: relaxin 1 #status predicted <MAT>

Alignment Scores:

Pred. No.:	0.00881	Length:	166
Score:	100.50	Matches:	41
Percent Similarity:	36.94%	Conservative:	17
Best Local Similarity:	26.11%	Mismatches:	42
Query Match:	15.83%	Indels:	57
DB:	2	Gaps:	7

US-09-518-842-1_COPY_76_417 (1-342) x S42783 (1-166)

```
Qy          1 GCAGAGCTGAGGGGATGTGGT---CCCCGATTTGGAAAACACTTGCTGTCATATTGCCCC 57
              |||::::::      ||||||         :::   |||      ||||||
Db         23 AlaGlnIleAlaIleCysGlyMetSerThrTrpSerLysArgSerLeuSerGlnGluAsp 42

Qy         58 ATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCC 117
                 |||:::               ||||||                          ||||||
Db         43 AlaProGln--------------ThrPro----------------------ArgPro 49

Qy        118 ---AAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCA 174
                 |||:::|||   :::      |||||||||   :::   :::
Db         50 ValAlaGluIleValProSerPheIleAsnLysAspThrGluThrIleIleIleMetLeu 69

Qy        175 GAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA 234
              |||||||||   ||||||   ||||||||||||      |||||||||  ||||||||||
Db         70 GluPheIleAlaAsnLeuProProGluLeuLysAlaAlaLeuSerGluArgGlnProSer 89

Qy        235 TTG---------------------------------------------------------- 237
              |||
Db         90 LeuProGluProGlnGlnTyrValProAlaLeuLysAspSerAsnLeuSerPheGluGlu 109

Qy        238 ---AAGAAAATAATACTTTCCCGCAAAAAGAGAAGT----------------------- 270
                 ||||||:::|||   :::|||:::      :::
Db        110 PheLysLysLeuIleArgAsnArgGlnSerGluAlaAlaAspSerAsnProSerGluLeu 129

Qy        271 ----------------------GGACGTCACAGATTTGATCCATTC------------ 294
                                     :::   :::      |||:::
Db        130 LysTyrLeuGlyLeuAspThrHisSerGlnLysLysArgGlnProTyrValAlaLeuPhe 149

Qy        295 ------TGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                    ||||||   :::   |||         :::            |||
Db        150 GluLysCysCysLeuIleGlyCysThrLysArgSerLeuAlaAsnTyrCys 166
```

RESULT 11
RXRT

relaxin precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change 24-Sep-1999
C;Accession: A01614
R;Hudson, P.; Haley, J.; Cronk, M.; Shine, J.; Niall, H.
Nature 291, 127-131, 1981
A;Title: Molecular cloning and characterization of cDNA sequences coding for rat rela
A;Reference number: A01614; MUID:81197624; PMID:7231533
A;Accession: A01614
A;Molecule type: mRNA
A;Residues: 1-186 <HUD>
A;Cross-references: GB:J00780; GB:M25468; NID:g206606; PIDN:AAA42029.1; PID:g206607
C;Superfamily: insulin
C;Keywords: hormone; ovary; pyroglutamic acid
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-57/Domain: relaxin chain B #status predicted <RXB>
F;23-57,163-186/Product: relaxin #status predicted <MAT>
F;58-162/Domain: relaxin connecting C peptide #status predicted <RXC>
F;163-186/Domain: relaxin chain A #status predicted <RXA>
F;36-173,48-186,172-177/Disulfide bonds: #status predicted
F;163/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predi

Alignment Scores:

Pred. No.:	0.0192	Length:	186
Score:	97.50	Matches:	38
Percent Similarity:	31.79%	Conservative:	10
Best Local Similarity:	25.17%	Mismatches:	48
Query Match:	15.35%	Indels:	55
DB:	1	Gaps:	4

US-09-518-842-1_COPY_76_417 (1-342) x RXRT (1-186)

```
Qy         16 TGTGGTCCCCGATTTGGAAAACACTTGCTGTCATAT---TGCCCCATGCCTGAGAAGACA 72
              ||||||         |||:::      ||||||         |||   |||
Db         48 CysGlyAlaSerValGlyArgLeuAlaLeuSerGlnGluGluProAlaPro--------- 64

Qy         73 TTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCA 132
                                          |||                  |||:::|||
Db         65 --------------------------LeuAlaArgGlnAlaThrAlaGluValValPro 75

Qy        133 ACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTG 192
                 :::      |||||||||   :::           |||   :::   :::||||||||||
Db         76 SerPheIleAsnLysAspAlaGluProPheAspMetThrLeuLysCysLeuProAsnLeu 95

Qy        193 TCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAG------------------------ 228
              |||   |||   |||      ||||||||||||:::
Db         96 SerGluGluArgLysAlaAlaLeuSerGluGlyArgAlaProPheProGluLeuGlnGln 115

Qy        228 ---------------------------------------------------------- 228

Db        116 HisAlaProAlaLeuSerAspSerValValSerLeuGluGlyPheLysLysThrPheHis 135

Qy        229 --------------------------------CCATCATTGAAGAAAATA--------- 246
                                               |||   ||||||   :::
Db        136 AsnGlnLeuGlyGluAlaGluAspGlyGlyProProGluLeuLysTyrLeuGlySerAsp 155

Qy        247 ATACTTTCCCGCAAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTA 306
                 ||||||||||||||||                                ||||||   :::
Db        156 AlaGlnSerArgLysLysArgGlnSerGlyAlaLeuLeuSerGluGlnCysCysHisIle 175

Qy        307 ATTTGTGACGATGGAACTTCAGTTAAATTATGT 339
                 |||         :::      |||||||||
Db        176 GlyCysThrArgArgSerIleAlaLysLeuCys 186
```

RESULT 12
E83259
hypothetical protein PA3089 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83259
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.

```
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83259
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-321 <STO>
A;Cross-references: GB:AE004733; GB:AE004091; NID:g9949194; PIDN:AAG06477.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA3089
C;Superfamily: Neisseria meningitidis hypothetical protein NMB1349

Alignment Scores:
Pred. No.:                0.827          Length:          321
Score:                    83.00          Matches:          17
Percent Similarity:       61.54%         Conservative:      7
Best Local Similarity:    43.59%         Mismatches:        6
Query Match:              13.07%         Indels:            9
DB:                       2              Gaps:              2

US-09-518-842-1_COPY_76_417 (1-342) x E83259 (1-321)

Qy         19 GGTCCCCGATTTGGAAAACACTTGCTGTCAT-----------------ATTGCCCCATG 60
              ||||||:::   |||:::|||   ||| |||                 :::||||||
Db        234 GlyProGlnThrGlyGlnHisTrpLeu-HisLeuSerArgGluHisTrpLeuAlaProAl 253

Qy         61 CCTGAGAAGACATTCACCA------CCACCCCAGGAGGGTGGCTGCTGGAATCTG 109
                :::||||||      |||      |||||:::::|||   |||||||||:::
Db        253 aMetArgArgGlnAspProLeuAspProProArgAspGlyAlaCysTrpAsnIle 271

RESULT 13
AE0614
probable competence-related protein STY0984 [imported] - Salmonella enterica subsp. ente
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AE0614
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; PMID:11677608
A;Accession: AE0614
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-754 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05383.1; PID:g16502146; GSPDB:GN00176
C;Genetics:
A;Gene: STY0984

Alignment Scores:
Pred. No.:                3.2            Length:          754
Score:                    78.00          Matches:          32
Percent Similarity:       40.65%         Conservative:     18
Best Local Similarity:    26.02%         Mismatches:       43
Query Match:              13.00%         Indels:           31
DB:                       2              Gaps:              6

US-09-518-842-1_COPY_76_417 (1-342) x AE0614 (1-754)

Qy        324 AGTTCCATCGTCACAAATTACTTCACAACAGAATGGATCAAATCTGTGACGTCCACTTCT 265
              :::::::::::::   |||   |||         |||:::      ||||||
Db        386 ThrAlaMetLeuAlaAsnLeuPheAlaValProTrpValThrPheValThrValProLeu 405

Qy        264 CTTTTTGCGGGAAAGTAT-----------------TATTTTCTTCAATGATGGCTGCCC 223
              :::   ||||||                     :::|||||||:::   |||:::
Db        406 IleLeuAlaGlyMetIleLeuHisLeuThrGlyProPhePheLeuGluGluTrpVal--- 424

Qy        222 TTCAGACAGTGGTTTCTT------------------------------------------ 205
                  |||:::|||
Db        425 ---------TrpTyrLeuThrAspArgAlaLeuAlaAlaLeuPheTyrLeuLeuAsnAla 441

Qy        204 ------CAGCTCTGGTGACAAATTAGGAATGAATTCTG----ATGTCGTACCTAAGGCTT 155
                   |||   |||       |||   |||   :::      ::::::   |||
Db        442 LeuProGlnGlyTrpValAsnIleAspAsnArgTrpGlnTrpLeuThrLeuLeuProTrp 461

Qy        154 GTCCATCTTTGTTGTTGGAGGTTGACACCATTTCTTTGGGACGTCCAGATTCCAGCAGCC 95
              :::   |||      |||||||||         :::|||             ||||||
Db        462 LeuThrLeuIleAlaTrpArgLeuAsn------ValTrpArgThrTrp---ProAlaVal 478

Qy         94 ACCCTCCTGGGGTGGTGGTGAATGTCTTCTCAGGCATGGGGCAATATGACAGCAAGTGTT 35
                 |||   |||         ||||||         |||      :::   ||||||
Db        479 CysLeuCysGlyLeuLeuLeuMetSerTrpProLeuTrpArgProIleAsnAlaSerGly 498

Qy         34 TTCCAAATC 26
              :::|||:::
Db        499 TrpGlnVal 501

RESULT 14
AG0433
proteinase (EC 3.4.21.-) [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG0433
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AG0433
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92795.1; PID:g15981488; GSPDB:GN00175
C;Genetics:
A;Gene: degQ
C;Superfamily: Helicobacter serine proteinase
C;Keywords: hydrolase; serine proteinase

Alignment Scores:
Pred. No.:                3.97           Length:          457
Score:                    77.00          Matches:          25
Percent Similarity:       47.89%         Conservative:      9
Best Local Similarity:    35.21%         Mismatches:       27
Query Match:              12.13%         Indels:           10
DB:                       2              Gaps:              3

US-09-518-842-1_COPY_76_417 (1-342) x AG0433 (1-457)

Qy         31 GGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCCCAGGA 90
              ||||||   :::   |||:::      :::      |||      ||||||   ||||||
Db        321 GlyLysLysIleSerSerPheAlaGluLeuArgAlaLysValGlyThrThrGlyProGly 340

Qy         91 GGG-----------TGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCC 138
                           ||||||      |||:::|||   |||:::      |||
Db        341 LysThrIleLysIleGlyLeuLeuArgGluGlyLysProLeuGluValSerValThrLeu 360

Qy        139 AACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCA 198
              :::|||                     :::||||||           ||||||||||||
Db        361 AspAsnSer--------------SerSerThrSerThrSerAlaGluAsnLeuSerPro 375

Qy        199 GAGCTGAAG---AAACCACTGTCTGAAGGGCAG 228
                 |||:::         ||||||   |||:::
Db        376 SerLeuGlnGlyAlaSerLeuSerAsnGlyGlu 386

RESULT 15
A49194
relaxin - guinea pig
C;Species: Cavia porcellus (guinea pig)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 16-Jul-1999
C;Accession: A49194
R;Lee, Y.A.; Bryant-Greenwood, G.D.; Mandel, M.; Greenwood, F.C.
```

Endocrinology 130, 1165-1172, 1992
A;Title: The complementary deoxyribonucleic acid sequence of guinea pig endometrial pror
A;Reference number: A49194; MUID:92164503; PMID:1537282
A;Accession: A49194
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-160 <LEE>
A;Cross-references: GB:S85964; NID:g246391; PIDN:AAB21586.1; PID:g246392
C;Superfamily: insulin

Alignment Scores:

Pred. No.:	4.14	Length:	160
Score:	76.50	Matches:	31
Percent Similarity:	35.09%	Conservative:	9
Best Local Similarity:	27.19%	Mismatches:	31
Query Match:	12.05%	Indels:	43
DB:	2	Gaps:	3

US-09-518-842-1_COPY_76_417 (1-342) x A49194 (1-160)

```
Qy         97 CTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAA 156
              :::|||  ||||||:::      |||::::::  :::|||  ||||||:::
Db         39 IleLeuGlySerGlyGlnSerAlaGluIleMetProSerSerIleAsnLysGluValAsp 58

Qy        157 GCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTG 216
              :::|||          |||   |||  ||||||     ||||||:::      |||
Db         59 SerLeuAsnMetLeu---GluSerIleAlaAsnLeuProGluGluLeuArgAlaMetLeu 77

Qy        217 TCTGAAGGGCAGCCATCA------------------------------------------ 234
                 |||   |||||||||
Db         78 ProGluLysGlnProSerSerProGlnLeuGlnGlnTyrValProAlaLeuLysAsnSer 97

Qy        235 -----------------TTGAAGAAAATAATACTTTCCCGC----------------- 258
                          |||   |||||||||      |||
Db         98 AsnValAlaValLysGluLeuAsnLysIleIleArgGlyArgGlnGluGluAlaGluAsp 117

Qy        259 ------------------------------------------------AAAAAGAGAAGT 270
                                                             |||||||||
Db        118 AsnSerHisSerLeuLeuLysAspPheAsnLeuAsnIleTyrSerProLysLysArgGln 137

Qy        271 GGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGT 312
                                  ||||||:::|||   |||
Db        138 LeuAspMetThrValSerGluLysCysCysGlnValGlyCys 151
```

Search completed: June 14, 2003, 19:46:49
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:35:45 ; Search time 35 Seconds
(without alignments)
4026.749 Million cell updates/sec

Title: US-09-518-842-1_COPY_76_417
Perfect score: 635
Sequence: 1 GCAGAGCTGAGGGGATGTGG.....CTTCAGTTAAATTATGTACA 342

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09518842/runat_14062003_175811_12270/app_query.fasta_1.519
-DB=SPTREMBL_21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09518842_@CGN_1_1_138_@runat_14062003_175811_12270 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	113.5	17.9	199	6	Q9N0T9	Q9n0t9 camelus dro
	2	104	16.4	180	6	Q9MYK8	Q9myk8 felis silve
	3	101	15.9	560	5	Q22328	Q22328 caenorhabdi
	4	83	13.1	321	16	Q9HZB9	Q9hzb9 pseudomonas
	5	82	12.9	795	11	Q9DBY0	Q9dby0 mus musculu
	6	80.5	12.7	889	10	Q9FHH9	Q9fhh9 arabidopsis
c	7	78	13.0	754	16	Q8ZQC3	Q8zqc3 salmonella
c	8	78	13.0	754	16	Q8Z802	Q8z802 salmonella
	9	77	12.1	457	16	Q8ZB58	Q8zb58 yersinia pe
c	10	75.5	12.6	718	6	Q9BE45	Q9be45 bos taurus
c	11	75.5	12.6	793	4	Q9H5Z2	Q9h5z2 homo sapien
	12	74.5	11.7	752	10	Q94H57	Q94h57 oryza sativ
	13	74	11.7	475	10	Q9FTX0	Q9ftx0 oryza sativ
	14	74	11.7	502	10	Q942M3	Q942m3 oryza sativ
	15	73.5	11.6	273	16	Q99SX5	Q99sx5 staphylococ
c	16	73.5	12.2	402	10	O82202	O82202 arabidopsis
	17	73	11.5	590	4	Q96G63	Q96g63 homo sapien
	18	73	11.5	636	4	Q9H9B6	Q9h9b6 homo sapien
	19	73	11.5	695	4	Q9H062	Q9h062 homo sapien
	20	73	11.5	747	4	Q96AE5	Q96ae5 homo sapien
	21	73	11.5	747	4	Q9NVG7	Q9nvg7 homo sapien
	22	73	11.5	1435	3	Q03291	Q03291 saccharomyc
	23	72.5	11.4	1432	5	Q95ZV4	Q95zv4 caenorhabdi
	24	72.5	11.4	1434	5	Q8T3E4	Q8t3e4 caenorhabdi
	25	72	11.3	391	2	Q44032	Q44032 ralstonia m
	26	72	11.3	483	16	Q8YBN0	Q8ybn0 brucella me
c	27	72	12.0	506	17	Q96XF4	Q96xf4 sulfolobus
	28	72	11.3	752	3	Q9URT2	Q9urt2 schizosacch
c	29	71.5	11.9	1217	5	P91457	P91457 caenorhabdi
	30	71.5	11.3	1530	2	Q93QE8	Q93qe8 chlamydia t
	31	71	11.2	204	4	Q9BRP8	Q9brp8 homo sapien
c	32	71	11.8	206	3	Q99395	Q99395 saccharomyc
	33	71	11.2	327	5	Q9W0Z8	Q9w0z8 drosophila
	34	71	11.2	697	5	Q9VU11	Q9vul1 drosophila
	35	71	11.2	836	12	Q66400	Q66400 duck hepati
	36	71	11.2	1417	12	Q89241	Q89241 western equ
	37	71	11.2	1490	12	Q88791	Q88791 western equ
	38	70.5	11.1	347	3	P87229	P87229 schizosacch
	39	70.5	11.1	476	5	Q961W9	Q961w9 drosophila
c	40	70.5	11.8	766	10	Q9LNC4	Q9lnc4 arabidopsis
	41	70.5	11.1	1037	3	Q8WZX1	Q8wzx1 neurospora
c	42	70.5	11.8	2015	5	Q9VAS6	Q9vas6 drosophila
	43	70	11.0	542	5	Q9NJD9	Q9njd9 onchocerca
c	44	70	11.7	693	4	Q9NX82	Q9nx82 homo sapien
c	45	70	11.7	747	4	Q9Y6A1	Q9y6a1 homo sapien

ALIGNMENTS

RESULT 1
Q9N0T9
ID Q9N0T9 PRELIMINARY; PRT; 199 AA.
AC Q9N0T9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Preprorelaxin.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20191540; PubMed=10727251;
RA Hombach-Klonisch S., Abd-Elnaeim M., Skidmore J.A., Leiser R.,
RA Fischer B., Klonisch T.;
RT "Ruminant relaxin in the pregnant one-humped camel.";
RL Biol. Reprod. 62:839-846(2000).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF254739; AAF67741.1; -.
DR HSSP; P01348; 1RLX.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.

```
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
SQ   SEQUENCE   199 AA;  22386 MW;  F69BBD7E4F4A8089 CRC64;

Alignment Scores:
Pred. No.:              0.000149      Length:         199
Score:                  113.50        Matches:        42
Percent Similarity:     30.95%        Conservative:   10
Best Local Similarity:  25.00%        Mismatches:     59
Query Match:            17.87%        Indels:         57
DB:                     6             Gaps:           4

US-09-518-842-1_COPY_76_417 (1-342) x Q9N0T9 (1-199)

Qy          7 CTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGC------------ 54
              ::::::   ||||||            :::      :::      |||
Db         32 ValLysAlaCysGlyArgGluLeuValArgLeuTrpIleGluIleCysGlySerValSer 51

Qy         55 ---------CCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTG------ 99
                    |||  |||                          |||         |||
Db         52 TrpGlyArgProAlaProArgProAlaProArgProAlaProLysProAlaLeuArgPro 71

Qy        100 --------------------------CTGGAATCTGGACGTCCCAAAGAAATGGTGTCA 132
                                         |||   ||||||   |||   |||::::::
Db         72 AlaLeuSerGlnAspLysLysProArgLeuArgSerGlyProProAlaGluIleMetPro 91

Qy        133 ACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTG 192
              :::|||      ||||||   :::   |||   |||      ||||||   |||||||||
Db         92 SerSerIleThrLysAspAlaGluThrLeuThrThrMetLeuGluPheThrProAsnLeu 111

Qy        193 TCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCA------------------ 234
                 ||||||         |||||||||   |||||||||
Db        112 ProGlnGluLeuThrAlaThrLeuSerGluArgGlnProSerAlaGluProGlnGlnPro 131

Qy        235 --------------------------------TTGAAGAAAATAATACTTTCCCGC--- 258
                                              ||||||||||||      |||
Db        132 AlaLeuLysAspSerAsnLeuAsnPheGluGluPheLysLysIleIlePheAspArgGln 151

Qy        258 ------------------------------------------------------------ 258

Db        152 AsnGluGluGluAspGluSerLeuSerGluLeuLysAsnLeuGlyLeuAspLysHisSer 171

Qy        259 ---AAAAAGAGAAGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGTGAC 315
                 |||||||||      :::                  |||||||:::      |||
Db        172 GluLysLysArgGlnLeuGlnMetThrLeuGlyGluArgCysCysGlnLysGlyCysSer 191

Qy        316 GATGGAACTTCAGTTAAATTATGT 339
                               |||
Db        192 ArgLysGluMetAlaThrAlaCys 199

RESULT 2
Q9MYK8
ID   Q9MYK8      PRELIMINARY;      PRT;   180 AA.
AC   Q9MYK8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Relaxin.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99115216; PubMed=9915995;
RA   Klonisch T., Hombach-Klonisch S., Froehlich C., Kauffold J.,
RA   Steger K., Huppertz B., Fischer B.;
RT   "Nucleic acid sequence of feline preprorelaxin and its localization
RT   within the feline placenta.";
RL   Biol. Reprod. 60:305-311(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hombach-Klonisch S., Klonisch T.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR   EMBL; AF233688; AAF60303.1; -.
DR   HSSP; P04090; 6RLX.
DR   InterPro; IPR004825; Ins/IGF/relax.
DR   Pfam; PF00049; Insulin; 1.
DR   SMART; SM00078; IlGF; 1.
DR   PROSITE; PS00262; INSULIN; 1.
SQ   SEQUENCE   180 AA;  20360 MW;  4C2CF371C698AF9F CRC64;

Alignment Scores:
Pred. No.:              0.00194       Length:         180
Score:                  104.00        Matches:        40
Percent Similarity:     32.72%        Conservative:   13
Best Local Similarity:  24.69%        Mismatches:     47
Query Match:            16.38%        Indels:         62
DB:                     6             Gaps:           4

US-09-518-842-1_COPY_76_417 (1-342) x Q9MYK8 (1-180)

Qy          7 CTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCCATG------ 60
              |||:::   ||||||      |||   :::      :::      |||
Db         30 LeuLysAlaCysGlyArgGluPheValArgLeuGlnIleArgIleCysGlySerLeuSer 49

Qy         61 ----------------------------CCTGAGAAGACATTCACCACCACCCCAGGA 90
                                           |||   :::                |||
Db         50 TrpGlyLysSerSerGlnGlnHisArgGluProArgGlnAlaProAlaAlaLeuPro--- 68

Qy         91 GGGTGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGAT 150
                                           |||:::||||||:::|||
Db         69 ----------------------------GluIleValSerSerSerIleThrSerGly 78

Qy        151 GGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAA 210
               :::||||||              |||:::||||||:::|||      |||||||||
Db         79 AlaGluAlaLeuAsnGlyMetLeuGluTyrIleProAspLeuProGlnGluLeuLysAla 98

Qy        211 CCACTGTCTGAAGGGCAGCCATCA---------------------------------- 234
                 |||||||||   :::||||||
Db         99 ThrLeuSerGluArgGluProSerPheArgGluLeuGlnProSerLeuLysAspSerAsn 118

Qy        235 ---------------TTGAAGAAAATAATACTT-------------------------- 252
                            :::::|||   ||||||
Db        119 LeuAsnLeuGluGluValGluLysSerIleLeuGlyArgGlnAsnGluAlaGluAspGln 138

Qy        253 ---------------------------------------TCCCGCAAAAAGAGAAGTGGA 273
                                                    ||||||   |||||||||
Db        139 SerLeuSerGlnLeuGlyArgSerArgLeuAspAlaHisSerArgIleLysArgSerAsp 158

Qy        274 CGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCAGTTAAA 333
                 |||:::         ||||||   |||   |||
Db        159 TyrIleArgTyrSerAspArgCysCysAsnValGlyCysThrArgLysGluLeuAlaAsp 178

Qy        334 TTATGT 339
                 ||||||
Db        179 LeuCys 180

RESULT 3
Q22328
ID   Q22328      PRELIMINARY;      PRT;   560 AA.
AC   Q22328;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 61.6 kDa protein.
GN   T07H6.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
```

```
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Geisel C.;
RT   "The sequence of C. elegans cosmid T07H6.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U53344; AAA96225.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   SMART; SM00032; CCP; 7.
KW   Hypothetical protein.
SQ   SEQUENCE   560 AA;  61619 MW;  FB8923BAC1B320C9 CRC64;

Alignment Scores:
Pred. No.:                0.00564        Length:          560
Score:                    101.00         Matches:         43
Percent Similarity:       38.56%         Conservative:    16
Best Local Similarity:    28.10%         Mismatches:      38
Query Match:              15.91%         Indels:          56
DB:                       5              Gaps:            9

US-09-518-842-1_COPY_76_417 (1-342) x Q22328 (1-560)

Qy          19 GGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACC 78
               ||||||         |||   |||   |||:::|||      |||   ||||||:::
Db         414 GlyProProMet---LysCysLeuProSerTrpCysGluHisProSerLysThrTyrGly 432

Qy          79 ACCACCCCAGGAGGGTGGCTGCTG---------------------------------- 102
               |||   |||||||||   :::|||
Db         433 ThrLeuProGlyGlyGlnIleLeuLeuGluGlyIleLeuGlyAlaTyrGluPheGlnSer 452

Qy         103 ---------------GAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 147
                             |||   ||||||          :::|||              |||
Db         453 TyrIleGlnLysValGluGluGlyArg---------AlaIleSerPheGlnCysGlyLys 469

Qy         148 GATGGACAAGCCTTAGGT-----------------ACGACATCAGAATTCATTCCTAAT 189
                        :::|||                                 |||::::::|||
Db         470 GlyAsnTyrLeuIleGlyProProLysAlaThrCysValAsnGlyGluTrpMetProLys 489

Qy         190 TTGTCACCAGAGCTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAG---AAAATA 246
               :::|||||||         |||   :::|||:::      |||   ::::::   ||||||
Db         490 ValSerPro---------LysCysValSerGlnThrHisProMetIleGluGlyLysIle 506

Qy         247 ATACTTTCCCCCAAAAAGAGAAGT------GGACGT----------------------- 276
               :::      ||||||||||||||      ||||||
Db         507 LeuTrpAspArgLysLysArgSerLeuProGlyArgAlaValArgGluTyrValAspAsp 526

Qy         277 ------------CACAGATTTGATCCATTCTGTTGTGAAGTAATT--------------- 309
                        ||||||                  |||   ::::::
Db         527 GluLeuSerThrHisArgGlnHisSerGlyLysCysGlyIleValSerGlyLysLeuGlu 546

Qy         310 -----------------TGTGACGATGGAACTTCAGTT 330
                               |||:::|||   ||||||
Db         547 ArgMetIleMetGlnHisSerAspAsnGlyValSerVal 559

RESULT 4
Q9HZB9
ID   Q9HZB9      PRELIMINARY;      PRT;   321 AA.
AC   Q9HZB9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Hypothetical protein PA3089.
GN   PA3089.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004733; AAG06477.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   321 AA;  36640 MW;  3EE85FDD3DBF268F CRC64;

Alignment Scores:
Pred. No.:                0.671          Length:          321
Score:                    83.00          Matches:         17
Percent Similarity:       61.54%         Conservative:    7
Best Local Similarity:    43.59%         Mismatches:      6
Query Match:              13.07%         Indels:          9
DB:                       16             Gaps:            2

US-09-518-842-1_COPY_76_417 (1-342) x Q9HZB9 (1-321)

Qy          19 GGTCCCCGATTTGGAAAACACTTGCTGTCAT-----------------ATTGCCCCATG 60
               ||||||:::   |||:::|||   ||| |||                 :::|||||||
Db         234 GlyProGlnThrGlyGlnHisTrpLeu-HisLeuSerArgGluHisTrpLeuAlaProAl 253

Qy          61 CCTGAGAAGACATTCACCA------CCACCCCAGGAGGGTGGCTGCTGGAATCTG 109
                  :::||||||       |||      ||||||:::::::|||   ||||||||||:::
Db         253 aMetArgArgGlnAspProLeuAspProProArgAspGlyAlaCysTrpAsnIle 271

RESULT 5
Q9DBY0
ID   Q9DBY0      PRELIMINARY;      PRT;   795 AA.
AC   Q9DBY0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   1200010K03Rik protein.
GN   1200010K03RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LUNG;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
```

```
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004693; BAB23479.1; -.
DR   HSSP; Q63245; 2HFH.
DR   MGD; MGI:1921373; 1200010K03Rik.
DR   InterPro; IPR001766; TF_Fork_head.
DR   InterPro; IPR000822; Znf_C2H2.
DR   Pfam; PF00250; Fork_head; 1.
DR   Pfam; PF00096; zf-C2H2; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
SQ   SEQUENCE   795 AA;  85980 MW;  860AE60AAD3742CE CRC64;

Alignment Scores:
Pred. No.:              1.08          Length:        795
Score:                  82.00         Matches:       33
Percent Similarity:     40.87%        Conservative:  14
Best Local Similarity:  28.70%        Mismatches:    36
Query Match:            12.91%        Indels:        32
DB:                     11            Gaps:          9

US-09-518-842-1_COPY_76_417 (1-342) x Q9DBY0 (1-795)

Qy         52 TGCCCCATGCCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGA 111
              ||||||                ||| |||  |||  |||  |||  |||
Db        228 CysPro-----------------ThrAspLeuProGlnLeuTrpLysGlyGlu---Gly 240

Qy        112 CGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACA 171
                 |||   :::          |||      :::::::|||      |||   :::|||
Db        241 AlaProGlyGlnProAlaGluAspSerGlyArgGlnGluGlyLeuAspLeuAlaSerThr 260

Qy        172 ------------TCAGAATTCATT------CCTAATTTGTCACCAGAGCTGAAG---AAA 210
                         :::   |||         |||   :::||||||   |||
Db        261 AlaValThrAlaThrSerPheAlaSerProProLysValSerProProLeuSerHisHis 280

Qy        211 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 270
              ||||||     ||||||||||:::            :::|||:::   ::::::   |||
Db        281 ProLeuProAsnGlyGlnProThr------------ValLeuThrSerArgArgAspSer 296

Qy        271 GGACGTCACAGATTTGATCCATTC--------------------------TGT------ 297
                 |||         |||                              |||
Db        297 SerSerHisGluGluThrProSerSerHisProLeuTyrGlyHisGlyGluCysLysTrp 316

Qy        298 ------TGTGAAGTAATTTGTGACGAT---GGAACTTCAGTTAAA 333
                    ||||||   :::|||:::|||   |||      :::|||
Db        317 ProGlyCysGluThrLeuCysGluAspLeuGlyGlnPheIleLys 331

RESULT 6
Q9FHH9
ID   Q9FHH9      PRELIMINARY;      PRT;   889 AA.
AC   Q9FHH9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   DNA-binding protein-like.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20181125; PubMed=10718197;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63(2000).
DR   EMBL; AB018117; BAB11612.1; -.
DR   EMBL; AB025628; BAB11612.1; JOINED.
DR   InterPro; IPR001878; Znf_CCHC.
DR   InterPro; IPR001841; Znf_ring.
DR   SMART; SM00184; RING; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
KW   DNA-binding.
SQ   SEQUENCE   889 AA;  98650 MW;  C616ACF664CC0CCE CRC64;

Alignment Scores:
Pred. No.:              1.66          Length:        889
Score:                  80.50         Matches:       29
Percent Similarity:     38.71%        Conservative:  19
Best Local Similarity:  23.39%        Mismatches:    43
Query Match:            12.68%        Indels:        33
DB:                     10            Gaps:          5

US-09-518-842-1_COPY_76_417 (1-342) x Q9FHH9 (1-889)

Qy         37 CACTTGCTGTCATATTGCCCC---------------------------------------- 57
                 |||   :::   :::|||||||
Db        214 HisPheIleGlnHisCysProThrAsnGlyAspProAsnTyrAspValLysArgValLys 233

Qy         58 -----------ATGCCTGAGAAGACATTCACCACCACCCCA---GGAGGGTGGCTGCTG 102
                        :::|||:::                |||||   |||   :::   |||
Db        234 ProProThrGlyIleProLysSerMetLeuMetAlaThrProAspGlySerTyrSerLeu 253

Qy        103 GAATCTGGA--------------CGTCCCAAAGAAATGGTGTCAACCTCCAACAACAAA 147
                 |||||               :::|||   |||           :::      |||
Db        254 ProSerGlyAlaValAlaValLeuLysProAsnGlu---------AspAlaPheGluLys 270

Qy        148 GATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAG 207
                 :::   :::   |||   :::|||:::      :::      |||   ||||||||||||
Db        271 GluMetGluGlyLeuProSerThrThrArgSerValGlyGluLeuProProGluLeuLys 290

Qy        208 AAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGA 267
                 ||||||                        |||   ::::::
Db        291 CysProLeuCys--------------------LysGluValMetLysAspAlaAlaLeu 303

Qy        268 AGTGGACGTCACAGATTTGATCCATTCTGTTGTGAAGTAATTTGTGACGATGGAACTTCA 327
                 :::   :::      :::      ||||||   :::   |||   |||         |||
Db        304 ThrSerLysCysCysTyrLysSerPheCysAspLysCysIleArgAspHisIleIleSer 323

Qy        328 GTTAAATTATGT 339
                 :::|||
Db        324 LysSerMetCys 327

RESULT 7
Q8ZQC3
ID   Q8ZQC3      PRELIMINARY;      PRT;   754 AA.
AC   Q8ZQC3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative recombination protein.
GN   YCAI OR STM0983.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
```

```
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008742; AAL19917.1; -.
DR   InterPro; IPR001279; Blactmase-like.
DR   InterPro; IPR004477; ComEC_N-term.
DR   InterPro; IPR004797; ComEC_Rec2.
DR   Pfam; PF00753; lactamase_B; 1.
DR   TIGRFAMs; TIGR00360; ComEC_N-term; 1.
DR   TIGRFAMs; TIGR00361; ComEC_Rec2; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   754 AA;  84954 MW;  711A95D282271358 CRC64;

Alignment Scores:
Pred. No.:              3.16          Length:        754
Score:                  78.00         Matches:       32
Percent Similarity:     40.65%        Conservative:  18
Best Local Similarity:  26.02%        Mismatches:    43
Query Match:            13.00%        Indels:        31
DB:                     16            Gaps:          6

US-09-518-842-1_COPY_76_417 (1-342) x Q8ZQC3 (1-754)

Qy        324 AGTTCCATCGTCACAAATTACTTCACAACAGAATGGATCAAATCTGTGACGTCCACTTCT 265
              :::::::::::   |||   |||         |||:::      ||||||
Db        386 ThrAlaMetLeuAlaAsnLeuPheAlaValProTrpValThrPheValThrValProLeu 405

Qy        264 CTTTTTGCGGGAAAGTAT----------------TATTTTCTTCAATGATGGCTGCCC 223
              :::   ||||||                    :::|||||||:::   |||:::
Db        406 IleLeuAlaGlyMetIleLeuHisLeuThrGlyProPhePheLeuGluGluTrpVal--- 424

Qy        222 TTCAGACAGTGGTTTCTT------------------------------------------ 205
                  |||:::|||
Db        425 ---------TrpTyrLeuThrAspArgAlaLeuAlaAlaLeuPheTyrLeuLeuAsnAla 441

Qy        204 ------CAGCTCTGGTGACAAATTAGGAATGAATTCTG----ATGTCGTACCTAAGGCTT 155
                  |||   |||      |||   |||   :::      ::::::   |||
Db        442 LeuProGlnGlyTrpValAsnIleAspAsnArgTrpGlnTrpLeuThrLeuLeuProTrp 461

Qy        154 GTCCATCTTTGTTGTTGGAGGTTGACACCATTTCTTTGGGACGTCCAGATTCCAGCAGCC 95
              :::   |||      |||||||||         :::|||           ||||||
Db        462 LeuThrLeuIleAlaTrpArgLeuAsn------ValTrpArgThrTrp---ProAlaVal 478

Qy         94 ACCCTCCTGGGGGTGGTGGTGAATGTCTTCTCAGGCATGGGGCAATATGACAGCAAGTGTT 35
                 |||   |||         ||||||         |||      :::   ||||||
Db        479 CysLeuCysGlyLeuLeuLeuMetSerTrpProLeuTrpArgProIleAsnAlaSerGly 498

Qy         34 TTCCAAATC 26
              :::|||:::
Db        499 TrpGlnVal 501

RESULT 8
Q8Z802
ID   Q8Z802      PRELIMINARY;      PRT;   754 AA.
AC   Q8Z802;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative competence-related protein.
GN   STY0984.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627268; CAD05383.1; -.
DR   InterPro; IPR001279; Blactmase-like.
DR   InterPro; IPR004477; ComEC_N-term.
DR   InterPro; IPR004797; ComEC_Rec2.
DR   Pfam; PF00753; lactamase_B; 1.
DR   TIGRFAMs; TIGR00360; ComEC_N-term; 1.
DR   TIGRFAMs; TIGR00361; ComEC_Rec2; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   754 AA;  84920 MW;  6697C033FB29576E CRC64;

Alignment Scores:
Pred. No.:              3.16          Length:        754
Score:                  78.00         Matches:       32
Percent Similarity:     40.65%        Conservative:  18
Best Local Similarity:  26.02%        Mismatches:    43
Query Match:            13.00%        Indels:        31
DB:                     16            Gaps:          6

US-09-518-842-1_COPY_76_417 (1-342) x Q8Z802 (1-754)

Qy        324 AGTTCCATCGTCACAAATTACTTCACAACAGAATGGATCAAATCTGTGACGTCCACTTCT 265
              :::::::::::   |||   |||         |||:::      ||||||
Db        386 ThrAlaMetLeuAlaAsnLeuPheAlaValProTrpValThrPheValThrValProLeu 405

Qy        264 CTTTTTGCGGGAAAGTAT----------------TATTTTCTTCAATGATGGCTGCCC 223
              :::   ||||||                    :::|||||||:::   |||:::
Db        406 IleLeuAlaGlyMetIleLeuHisLeuThrGlyProPhePheLeuGluGluTrpVal--- 424

Qy        222 TTCAGACAGTGGTTTCTT------------------------------------------ 205
                  |||:::|||
Db        425 ---------TrpTyrLeuThrAspArgAlaLeuAlaAlaLeuPheTyrLeuLeuAsnAla 441

Qy        204 ------CAGCTCTGGTGACAAATTAGGAATGAATTCTG----ATGTCGTACCTAAGGCTT 155
                  |||   |||      |||   |||   :::      ::::::   |||
Db        442 LeuProGlnGlyTrpValAsnIleAspAsnArgTrpGlnTrpLeuThrLeuLeuProTrp 461

Qy        154 GTCCATCTTTGTTGTTGGAGGTTGACACCATTTCTTTGGGACGTCCAGATTCCAGCAGCC 95
              :::   |||      |||||||||         :::|||           ||||||
Db        462 LeuThrLeuIleAlaTrpArgLeuAsn------ValTrpArgThrTrp---ProAlaVal 478

Qy         94 ACCCTCCTGGGGGTGGTGGTGAATGTCTTCTCAGGCATGGGGCAATATGACAGCAAGTGTT 35
                 |||   |||         ||||||         |||      :::   ||||||
Db        479 CysLeuCysGlyLeuLeuLeuMetSerTrpProLeuTrpArgProIleAsnAlaSerGly 498

Qy         34 TTCCAAATC 26
              :::|||:::
Db        499 TrpGlnVal 501

RESULT 9
Q8ZB58
ID   Q8ZB58      PRELIMINARY;      PRT;   457 AA.
AC   Q8ZB58;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Protease (EC 3.4.21.-).
GN   DEGQ OR YPO3566.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
```

```
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414157; CAC92795.1; -.
DR   MEROPS; S01.274; -.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001940; Protease2C.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00595; PDZ; 2.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00834; PROTEASES2C.
DR   SMART; SM00228; PDZ; 2.
DR   PROSITE; PS50106; PDZ; 2.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   457 AA;  47399 MW;  B89E79B8229FC86F CRC64;

Alignment Scores:
Pred. No.:                3.72          Length:         457
Score:                    77.00         Matches:        25
Percent Similarity:       47.89%        Conservative:   9
Best Local Similarity:    35.21%        Mismatches:     27
Query Match:              12.13%        Indels:         10
DB:                       16            Gaps:           3

US-09-518-842-1_COPY_76_417 (1-342) x Q8ZB58 (1-457)

Qy          31 GGAAAACACTTGCTGTCATATTGCCCCATGCCTGAGAAGACATTCACCACCACCCCAGGA 90
               ||||||   :::   |||:::      :::      |||      ||||||   ||||||
Db         321 GlyLysLysIleSerSerPheAlaGluLeuArgAlaLysValGlyThrThrGlyProGly 340

Qy  -       91 GGG------------TGGCTGCTGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCC 138
                             ||||||      |||:::|||   |||:::      |||
Db         341 LysThrIleLysIleGlyLeuLeuArgGluGlyLysProLeuGluValSerValThrLeu 360

Qy         139 AACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCA 198
               :::|||                  :::||||||              ||||||||||||
Db  .      361 AspAsnSer---------------SerSerThrSerThrSerAlaGluAsnLeuSerPro 375

Qy         199 GAGCTGAAG---AAACCACTGTCTGAAGGGCAG 228
                  |||:::         ||||||   |||:::
Db         376 SerLeuGlnGlyAlaSerLeuSerAsnGlyGlu 386

RESULT 10
Q9BE45
ID   Q9BE45      PRELIMINARY;      PRT;   718 AA.
AC   Q9BE45;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MAIL.
GN   MAIL.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HOLSTEIN; TISSUE=BLOOD;
RA   Yamaji D., Kitamura H., Morimatsu M., Shiina T., Kanehira K.,
RA   Fujikura D., Saito M.;
RT   "Bos taurus mRNA for MAIL, complete cds.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB058410; BAB39767.1; -.
DR   HSSP; P25963; 1NFI.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; ank; 6.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 6.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Repeat.
SQ   SEQUENCE   718 AA;  78247 MW;  CB8D8F6A994A9E00 CRC64;

Alignment Scores:
Pred. No.:                6.18          Length:         718
Score:                    75.50         Matches:        30
Percent Similarity:       39.05%        Conservative:   11
Best Local Similarity:    28.57%        Mismatches:     31
Query Match:              12.58%        Indels:         33
DB:                       6             Gaps:           5

US-09-518-842-1_COPY_76_417 (1-342) x Q9BE45 (1-718)

Qy         272 CCACTTCTCTTTTTGCGGGAAAGTATTATTTTCTTCAATGATGGCTGCCCTTCAGACAGT 213
               |||   ||||||   ||||||                                ||||||   |||
Db         324 ProHisLeuPheGlyArgGluPro--------------GlnPheCysPro---AspGln 337

Qy         212 GGTTTC------TTCAGCTCTGGTGACAAATTAGGAATGAATTCTGATGTCGTACCTAAG 159
                  |||         |||      :::         |||      |||   |||:::
Db         338 SerPheAlaProLeuLeuSerAspProArgGlnSerGluAsnIleAlaValProProGln 357

Qy         158 GCTTGTCCATCTTTGTTGTTGGAGGTTGACACC--------------------------- 126
                     ||||||:::      ::::::||||||
Db         358 ThrAlaProSerValGlnGlnGlnIleAspThrHisLeuGlnAsnPheSerLeuMetPro 377

Qy         125 ------------ATTTCTTTGGGACGTCCAGATTCCAGCAGCCACCCTCCTGGGGTGGTG 78
                           :::|||   |||||||||:::||||||   |||
Db         378 ProAspThrCysGluAlaLeuAlaArgProAspAlaSerSerThrProLeuSerThrPro 397

Qy          77 GTGAATGTCTTCTCAGGCATGGGGCAATATGACAGCAAGTGTTTTCCAAATCGGGGACCA 18
               :::                                              |||||||||   |||
Db         398 LeuPro-----------------------------------PheProAsnLeuGlyGly 405

Qy          17 CATCCCCTCAGCTCT 3
               :::|||:::|||:::
Db         406 AsnProMetSerThr 410

RESULT 11
Q9H5Z2
ID   Q9H5Z2      PRELIMINARY;      PRT;   793 AA.
AC   Q9H5Z2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CDNA: FLJ22757 fis, clone KAIA0803.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ILEAL MUCOSA;
RA   Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA   Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA   Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA   Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA   Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK026410; BAB15475.1; -.
DR   InterPro; IPR005112; dDENN.
DR   InterPro; IPR001194; DENN.
DR   InterPro; IPR005113; uDENN.
DR   Pfam; PF03455; dDENN; 1.
DR   Pfam; PF02141; DENN; 1.
DR   Pfam; PF03456; uDENN; 1.
SQ   SEQUENCE   793 AA;  86230 MW;  3F37ACA836F24AF4 CRC64;

Alignment Scores:
Pred. No.:                6.32          Length:         793
```

```
Score:                 75.50          Matches:       26
Percent Similarity:    42.11%         Conservative:  14
Best Local Similarity: 27.37%         Mismatches:    26
Query Match:           12.58%         Indels:        29
DB:                    4              Gaps:          4

US-09-518-842-1_COPY_76_417 (1-342) x Q9H5Z2 (1-793)

Qy        236 AATGATGGCTGCCCT----------------TCAGACAGTGGTTTCTTCAGCTCTGGTGAC 192
              ::::::||||||||||                  ||||||   |||      ||||||:::
Db        536 AspGluGlyCysProTrpAlaGluGluAlaLeuAspSerSerPheLeuGlySerGlyGlu 555

Qy        191 AAATTAGGAATGAATTCTGATGTCGTACCTAAGGCTTGTCCATCTTTGTTGTTGGAGGTT 132
              :::|||   :::   |||::::::                                 :::
Db        556 GluLeuAspLeuLeuSerGluIle--------------------------------Leu 564

Qy        131 GACACCATTTCTTTGGGACGTCCAGATTCCAGCAGCCACCCTCCT--------------- 87
              |||:::::::|||:::|||         :::   |||      |||
Db        565 AspSerLeuSerMetGlyAlaLysSerAlaGlySerLeuArgProSerGlnSerLeuAsp 584

Qy         86 ------------GGGGTGGTGGTGAATGTCTTCTCAGGCATGGGG------------CAA 51
                         |||   :::   :::   ||||||                      |||
Db        585 CysCysHisArgGlyAspLeuAspSerCysPheSerLeuProAsnIleLeuArgTrpGln 604

Qy         50 TATGACAGCAAGTGTTTTCCAAATCGGGGACCACATCCCCTCAGC 6
                 |||   |||      |||        |||   |||||||||
Db        605 ProAspAspLysLysLeuProGluProGluProGlnProLeuSer 619

RESULT 12
Q94H57
ID   Q94H57      PRELIMINARY;      PRT;   752 AA.
AC   Q94H57;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative retrotransposon protein.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,
RA   Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tsitrin T.,
RA   Riggs F., Hsiao J., Zismann V., Blunt S., Pai G., VanAken S.E.,
RA   Utterback T.R., Feldblyum T.V., Quackenbush J., Salzberg S.L.,
RA   White O., Fraser C.M.;
RT   "Oryza sativa chromosome 3 BAC OSJNBb0057P11 genomic sequence.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC084767; AAK72287.1; -.
SQ   SEQUENCE   752 AA;  83920 MW;  88AC2AEC1157DB98 CRC64;

Alignment Scores:
Pred. No.:             8.2            Length:        752
Score:                 74.50          Matches:       22
Percent Similarity:    48.61%         Conservative:  13
Best Local Similarity: 30.56%         Mismatches:    34
Query Match:           11.73%         Indels:        4
DB:                    10             Gaps:          2

US-09-518-842-1_COPY_76_417 (1-342) x Q94H57 (1-752)

Qy         47 CATATTGCCCCATGCCTG------AGAAGACATTCACCACCACCCCAGGAGGGTGGCTGC 100
              |||:::   |||   :::              :::||||||||||   |||
Db        280 HisLeuLeuProValValProTyrAsnAlaPheAsnProProProProGluPheSerLeu 299

Qy        101 TGGAATCTGGACGTCCCAAAGAAATGGTGTCAACCTCCAAC-AACAAAGATGGACAAGCC 159
                  |||   |||          |||:::   |||      :::   |||||||||
Db        300 MetLysSerAspProProThrAlaGlnCysArgSerProArgArgGlnThrGlyGlnAla 319

Qy        160 TTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAGCTGAAGAAACCACTGTCT 219
                    |||   :::   :::|||   |||   :::   |||         ||||||   :::
Db        320 SerGly---SerProLysPheGlnProThrValGlnProSerSerProLysProThrAsn 338

Qy        220 GAAGGGCAGCCATCATTGAAGAAAATAATACTTTCC 255
                 :::               :::|||::::::||||||
Db        339 GlnSerGlySerGlnLysArgLysLeuValLeuSer 350

RESULT 13
Q9FTX0
ID   Q9FTX0      PRELIMINARY;      PRT;   475 AA.
AC   Q9FTX0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   P0407B12.10 protein.
GN   P0407B12.10.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0407B12.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002843; BAB17173.1; -.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
SQ   SEQUENCE   475 AA;  53631 MW;  A9E4A62BF6AB21B3 CRC64;

Alignment Scores:
Pred. No.:             8.49           Length:        475
Score:                 74.00          Matches:       28
Percent Similarity:    44.34%         Conservative:  19
Best Local Similarity: 26.42%         Mismatches:    45
Query Match:           11.65%         Indels:        14
DB:                    10             Gaps:          3

US-09-518-842-1_COPY_76_417 (1-342) x Q9FTX0 (1-475)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              :::|||||||||           :::|||   |||||||||           |||:::
Db        269 SerGluLeuArgAsnArgAlaLeuGlnPheValLysHisLeuGlyLysThrLeuProLeu 288

Qy         61 CCTGAGAAGACATTCACCACC-------------------------ACCCCAGGAGGG 93
                         :::|||                             |||
Db        289 ThrSerSerAsnLeuSerThrHisHisHisLeuLysLeuIleTyrValValProAlaPro 308

Qy         94 TGGCTGCTGGAATCTGGA-----------CGTCCCAAAGAAATGGTGTCAACCTCCAAC 141
                 :::   |||                |||||:::::::::::::      ::::::
Db        309 ArgMetLysGluLeuAsnAsnGlnGluPheArgProGluAspLeuLeuGlnLeuAlaAsp 328

Qy        142 AACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAG 201
              :::   ||||||   :::|||      |||   :::|||            :::|||
Db        329 SerValAspGlyPheSerLeu---MetThrTyrAspPheSerGlyProGlnAsnProGly 347

Qy        202 CTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAA 261
                  |||||||||       |||   |||||||||   ::::::   ::::::
Db        348 ProSerAlaProLeuSerTrpIleGlnTyrSerLeuLysThrLeuLeuAlaAlaLysGly 367

Qy        262 AAGAGAAGTGGACGTCAC 279
                 |||      |||
Db        368 SerAlaSerHisGlyHis 373

RESULT 14
Q942M3
ID   Q942M3      PRELIMINARY;      PRT;   502 AA.
AC   Q942M3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   B1040D09.22 protein.
GN   B1040D09.22.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:B1040D09.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003328; BAB67973.1; -.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
SQ   SEQUENCE   502 AA;  56877 MW;  0B6EA76DF4555660 CRC64;

Alignment Scores:
Pred. No.:              8.59          Length:         502
Score:                  74.00         Matches:        28
Percent Similarity:     44.34%        Conservative:   19
Best Local Similarity:  26.42%        Mismatches:     45
Query Match:            11.65%        Indels:         14
DB:                     10            Gaps:           3

US-09-518-842-1_COPY_76_417 (1-342) x Q942M3 (1-502)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              :::|||||||||            :::|||   ||||||||            |||::: 
Db        296 SerGluLeuArgAsnArgAlaLeuGlnPheValLysHisLeuGlyLysThrLeuProLeu 315

Qy         61 CCTGAGAAGACATTCACCACC-------------------------ACCCCAGGAGGG 93
                       :::|||                                 |||
Db        316 ThrSerSerAsnLeuSerThrHisHisHisLeuLysLeuIleTyrValValProAlaPro 335

Qy         94 TGGCTGCTGGAATCTGGA------------CGTCCCAAAGAAATGGTGTCAACCTCCAAC 141
                 :::   |||                 ||||||:::::::::::::     ::::::
Db        336 ArgMetLysGluLeuAsnAsnGlnGluPheArgProGluAspLeuLeuGlnLeuAlaAsp 355

Qy        142 AACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTCATTCCTAATTTGTCACCAGAG 201
              :::   ||||||   :::|||      |||   :::|||            :::|||
Db        356 SerValAspGlyPheSerLeu---MetThrTyrAspPheSerGlyProGlnAsnProGly 374

Qy        202 CTGAAGAAACCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAA 261
                    |||||||||      |||   ||||||||   ::::::   ::::::
Db        375 ProSerAlaProLeuSerTrpIleGlnTyrSerLeuLysThrLeuLeuAlaAlaLysGly 394

Qy        262 AAGAGAAGTGGACGTCAC 279
                  |||     |||
Db        395 SerAlaSerHisGlyHis 400

RESULT 15
Q99SX5
ID   Q99SX5      PRELIMINARY;      PRT;   273 AA.
AC   Q99SX5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   NAD synthetase, prefers NH3 over glutamine.
GN   NADE OR SAV1912 OR SA1728.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003363; BAB58074.1; -.
DR   EMBL; AP003135; BAB42998.1; -.
DR   HSSP; P08164; 2NSY.
DR   InterPro; IPR003694; NAD_synthase.
DR   Pfam; PF02540; NAD_synthase; 1.
DR   TIGRFAMs; TIGR00552; nadE; 1.
KW   Complete proteome.
SQ   SEQUENCE   273 AA;  30683 MW;  EB4A022F38247F08 CRC64;

Alignment Scores:
Pred. No.:              8.62          Length:         273
Score:                  73.50         Matches:        30
Percent Similarity:     44.23%        Conservative:   16
Best Local Similarity:  28.85%        Mismatches:     30
Query Match:            11.57%        Indels:         28
DB:                     16            Gaps:           5

US-09-518-842-1_COPY_76_417 (1-342) x Q99SX5 (1-273)

Qy          1 GCAGAGCTGAGGGGATGTGGTCCCCGATTTGGAAAACACTTGCTGTCATATTGCCCCATG 60
              |||   :::   |||         |||   |||:::   ||||||:::|||
Db        181 AlaProIlePheGlyLeuAsnLysArgGlnGlyArgGlnLeuLeuAlaTyrLeuGlyAla 200

Qy         61 CCTGAGAAGACATTCACCACCACCCCAGGAGGGTGGCTGCTGGAATCTGGACGTCCCAAA 120
              |||::::::   :::      ||||||         ||||||      :::|||:::
Db        201 ProLysGluLeuTyrGluLysThrProThrAla---AspLeuGluAspAspLysProGln 219

Qy        121 GAAATGGTGTCAACCTCCAACAACAAAGATGGACAAGCCTTAGGTACGACATCAGAATTC 180
                                      |||      ||||||||   |||   |||
Db        220 --------------------LeuProAspGluAspAlaLeuGlyValThrTyrGluAla 232

Qy        181 ATTCCTAAT-----------------TTGTCACCAGAG-----------CTGAAGAAA 210
              |||   |||                 ::::::||||||           ::::::
Db        233 IleAspAsnTyrLeuGluGlyLysProValThrProGluGluGlnLysValIleGlu-As 252

Qy        211 CCACTGTCTGAAGGGCAGCCATCATTGAAGAAAATAATACTTTCCCGCAAAAAGAGAAGT 270
              ||||   ::::::   :::|||                              |||||||::
Db        252 nHisTyrIleArgAsnAlaHis-------------------------LysArgGluLe 263

Qy        271 GGACGTCACA 280
              :   |||
Db        263 uAlaTyrThr 266


Search completed: June 14, 2003, 19:46:07
Job time : 41 secs
```